IBM SPSS Bootstrapping 19



Note: Before using this information and the product it supports, read the general information under Notices auf S. 43.

Vorwort

IBM® SPSS® Statistics ist ein umfassendes System zum Analysieren von Daten. Das optionale Zusatzmodul Bootstrapping bietet die zusätzlichen Analyseverfahren, die in diesem Handbuch beschrieben sind. Die Prozeduren im Zusatzmodul Bootstrapping müssen zusammen mit SPSS Statistics Core verwendet werden. Sie sind vollständig in dieses System integriert.

Über SPSS Inc., ein Unternehmen von IBM

SPSS Inc., ein Unternehmen von IBM, ist ein führender globaler Anbieter von Analysesoftware und -lösungen zur Prognoseerstellung. Mit der vollständigen Produktpalette des Unternehmens – Datenerfassung, Statistik, Modellierung und Bereitstellung – werden Einstellungen und Meinungen von Personen erfasst und Ergebnisse von künftigen Interaktionen mit Kunden prognostiziert. Anschließend werden diese Erkenntnisse durch die Einbettung der Analysen in Geschäftsprozesse praktisch umgesetzt. Lösungen von SPSS Inc. sind durch die Konzentration auf die Zusammenführung von Analysefunktionen, IT-Architektur und Geschäftsprozessen für zusammenhängende unternehmensübergreifende Geschäftsziele konzipiert. Kunden aus den Bereichen Wirtschaft, Regierung und Wissenschaft vertrauen weltweit auf die Technologie von SPSS Inc. als Wettbewerbsvorteil, wenn es gilt, Kunden anzuziehen, zu binden und neue Kunden zu gewinnen und dabei Betrugsfälle zu verringern und Risiken zu entschärfen. SPSS Inc. wurde im Oktober 2009 von IBM übernommen. Weitere Informationen erhalten Sie unter *http://www.spss.com*.

Technischer Support

Kunden mit Wartungsvertrag können den Technischen Support in Anspruch nehmen. Kunden können sich an den Technischen Support wenden, wenn sie Hilfe bei der Arbeit mit den Produkten von SPSS Inc. oder bei der Installation in einer der unterstützten Hardware-Umgebungen benötigen. Wie Sie den Technischen Support kontaktieren können, entnehmen Sie der Website von SPSS Inc. unter *http://support.spss.com*. Über die Website unter *http://support.spss.com/default.asp?refpage=contactus.asp* können Sie auch nach Ihrem örtlichen Büro suchen. Wenn Sie Hilfe anfordern, halten Sie bitte Informationen bereit, um sich, Ihre Organisation und Ihren Supportvertrag zu identifizieren.

Kundendienst

Wenden Sie sich bei Fragen zur Lieferung oder Ihrem Kundenkonto an Ihr regionales Büro, das Sie auf der Website unter *http://www.spss.com/worldwide* finden. Halten Sie bitte stets Ihre Seriennummer bereit.

Ausbildungsseminare

SPSS Inc. bietet öffentliche und unternehmensinterne Seminare an. Alle Seminare beinhalten auch praktische Übungen. Seminare finden in größeren Städten regelmäßig statt. Wenn Sie weitere Informationen zu diesen Seminaren wünschen, wenden Sie sich an Ihr regionales Büro, das Sie auf der Website unter *http://www.spss.com/worldwide* finden.

Weitere Veröffentlichungen

Die Handbücher *SPSS Statistics: Guide to Data Analysis*, *SPSS Statistics: Statistical Procedures Companion* und *SPSS Statistics: Advanced Statistical Procedures Companion*, die von Marija Norušis geschrieben und von Prentice Hall veröffentlicht wurden, werden als Quelle für Zusatzinformationen empfohlen. Diese Veröffentlichungen enthalten statistische Verfahren in den Modulen “Statistics Base”, “Advanced Statistics” und “Regression” von SPSS. Diese Bücher werden Sie dabei unterstützen, die Funktionen und Möglichkeiten von IBM® SPSS® Statistics optimal zu nutzen. Dabei ist es unerheblich, ob Sie ein Neuling im Bereich der Datenanalyse sind oder bereits über umfangreiche Vorkenntnisse verfügen und damit in der Lage sind, auch die erweiterten Anwendungen zu nutzen. Weitere Informationen zu den Inhalten der Veröffentlichungen sowie Auszüge aus den Kapiteln finden Sie auf der folgenden Autoren-Website: *http://www.norusis.com*

Inhalt

Anhänge

Teil I: Benutzerhandbuch

Kapitel

1

Einführung in Bootstrapping

Bei der Erfassung von Daten sind Sie oft an den Eigenschaften der Grundgesamtheit interessiert, aus der Sie die Stichprobe genommen haben. Anhand von aus der Stichprobe berechneten Schätzwerten können Sie Schlussfolgerungen über diese Gesamtheitsparameter ziehen. Falls beispielsweise das im Lieferumfang des Produkts enthaltene Daten-Set *Employee data.sav* eine Zufallsstichprobe aus einer größeren Gesamtheit von Angestellten ist, ist der Stichprobenmittelwert von $34.419,57 für *Aktuelles Gehalt* eine Schätzung des durchschnittlichen aktuellen Gehalts für die Gesamtheit von Angestellten. Diese Schätzung hat zudem einen Standardfehler von $784,311 für eine Stichprobe der Größe 474, so dass $32.878,40 bis $35.960,73 ein 95%-Konfidenzintervall für das durchschnittliche aktuelle Gehalt in der Gesamtheit von Angestellten ist. Doch wie zuverlässig sind diese Schätzer? Für bestimmte "bekannte" Populationen und "well-behaved"-Parameter wissen wir einiges über die Eigenschaften der Stichprobenschätzer und können davon ausgehen, dass die Ergebnisse richtig sind. Bootstrapping dient dazu, mehr Informationen über die Eigenschaften von Schätzern für "unbekannte" Populationen und "ill-behaved"-Parameter zu gewinnen.

Abbildung 1-1
Ziehen von parametrischen Inferenzen über den Mittelwert der Grundgesamtheit

			Statistik	Standard-fehler
Gehalt	Mittelwert		$34,419.57	$784.311
	95%-Konfidenz-intervall	Untergrenze	$32,878.40	
		Obergrenze	$35,960.73	
	Median		$28,875.00	

Funktionsweise des Bootstrapping

Im einfachsten Fall nehmen Sie für ein Daten-Set mit einer Stichprobengröße *NB* "Bootstrap"-Stichproben der Größe *N* mit Zurücklegen aus dem ursprünglichen Datensatz und berechnen den Schätzer für jede dieser *B* Bootstrap-Stichproben. Diese *B* Bootstrap-Schätzungen sind eine Stichprobe der Größe *B* , anhand deren Sie Schlussfolgerungen über den Schätzer ziehen können. Nehmen Sie beispielsweise 1.000 Bootstrap-Stichproben aus dem Daten-Set *Mitarbeiterdaten.sav*, ist der anhand der Bootstraps geschätzte Standardfehler von $776,91 für den Stichprobenmittelwert von *Aktuelles Gehalt* eine Alternative zu dem Schätzwert von $784,311.

Des Weiteren bietet Bootstrapping einen Standardfehler und ein Konfidenzintervall für den Median, für den parametrische Schätzer nicht verfügbar sind.

Abbildung 1-2
Ziehen von Bootstrap-Inferenzen über den Mittelwert der Grundgesamtheit

			Statistik	Standard-fehler	Bootstrap[a]			
					Verzerrung	Standard-fehler	95%-Konfidenzintervall	
							Untergrenze	Obergrenze
Gehalt	Mittelwert		$34,419.57	$784.311	$14.66	$776.91	$32,990.38	$36,026.06
	95%-Konfidenz-intervall	Untergrenze	$32,878.40					
		Obergrenze	$35,960.73					
	Median		$28,875.00		$-13.22	$536.63	$27,750.00	$29,850.00

a.Unless otherwise noted, bootstrap results are based on 1000 bootstrap samples

Bootstrapping-Unterstützung innerhalb des Produkts

Bootstrapping ist bei Prozeduren, die es unterstützen, als untergeordnetes Dialogfeld enthalten. Weitere Informationen darüber, welche Prozeduren Bootstrapping unterstützen, finden Sie unter Prozeduren, die Bootstrapping unterstützen.

Wird Bootstrapping über das Dialogfeld angefordert, wird ein neuer und separater BOOTSTRAP-Befehl zusätzlich zu der üblichen, vom Dialogfeld generierten Syntax eingefügt. Durch den BOOTSTRAP–Befehl werden die Bootstrap-Stichproben gemäß Ihrer Spezifikationen erstellt. Intern werden diese Bootstrap-Stichproben als Aufteilungen behandelt, obwohl sie im Daten-Editor nicht explizit angezeigt werden. Dies bedeutet, dass es im Grunde genommen *B**N* Fälle gibt, weswegen die Anzeige in der Statusleiste im Laufe der Datenverarbeitung beim Bootstrapping von 1 bis *B**N* zählt. Das Ausgabeverwaltungssystem (OMS) wird verwendet, um die Ergebnisse zu erfassen, die durch die Ausführung der Analyse an jeder “Bootstrap-Aufteilung” gewonnen werden. Diese Ergebnisse werden gepoolt und zusammen mit den übrigen Ausgaben, die bei der Prozedur generiert wurden, im Viewer angezeigt. In bestimmten Fällen sehen Sie eine Referenz auf “bootstrap split 0”; dies ist das ursprüngliche Daten-Set.

Kapitel

2

Bootstrapping

Bootstrapping ist eine Methode zur Ableitung von robusten Schätzern von Standardfehlern und Konfidenzintervallen für Schätzer wie Mittel, Median, Anteil, Quotenverhältnis, Korrelationskoeffizient oder Regressionskoeffizient. Es kann auch für die Konstruktion von Hypothesentests verwendet werden. Bootstrapping ist besonders als Alternative zu parametrischen Schätzern geeignet, wenn die Annahmen dieser Methoden zweifelhaft (zum Beispiel bei Regressionsmodellen mit heteroskedastischen, auf kleine Stichproben angepassten Residuen) oder parametrische Schlussfolgerungen unmöglich sind oder äußerst komplizierte Formeln zur Berechnung von Standardfehlern erfordern (zum Beispiel bei der Berechnung von Konfidenzintervallen für den Median, Quartilen und andere Perzentilen).

Beispiele. Eine Telekommunikationsfirma verliert jeden Monat etwa 27 % ihrer Kunden durch Abwanderung. Um bei den Bemühungen zur Verringerung der Abwanderung die richtigen Schwerpunkte setzen zu können, möchte die Geschäftsleitung wissen, ob dieser Prozentsatz zwischen verschiedenen vordefinierten Kundengruppen variiert. Mit Bootstrapping können Sie ermitteln, ob sich die vier Hauptkundengruppen angemessen mit einer einzigen Abwanderungsquote beschreiben lassen. Für weitere Informationen siehe Thema Einsatz von Bootstrapping zum Berechnen von Konfidenzintervallen für Anteile in Kapitel 3 in *IBM SPSS Bootstrapping 19*.

Bei der Durchsicht von Mitarbeiterdaten ist die Geschäftsleitung an der bisherigen Arbeitserfahrung seiner Mitarbeiter interessiert. Die Arbeitserfahrung ist rechtslastig, was bedeutet, dass der Mittelwert eine weniger wünschenswerte Schätzung der “typischen” bisherigen Arbeitserfahrung unter Mitarbeitern darstellt als der Median. Parametrische Konfidenzintervalle sind allerdings für den Median im Produkt nicht enthalten. Für weitere Informationen siehe Thema Einsatz von Bootstrapping zum Berechnen von Konfidenzintervallen für Mediane in Kapitel 3 in *IBM SPSS Bootstrapping 19*.

Das Management ist ebenfalls daran interessiert zu ermitteln, welche Faktoren Gehaltserhöhungen von Mitarbeitern entsprechen, indem ein lineares Modell über die Differenz zwischen aktuellem und Anfangsgehalt erstellt wird. Beim Bootstrapping eines linearen Modells können Sie spezielle Resampling-Methoden (Residuen- und Wild-Bootstrap) verwenden, um genauere Ergebnisse zu erzielen. Für weitere Informationen siehe Thema Einsatz von Boostrapping zur Auswahl besserer Einflussvariablen in Kapitel 3 in *IBM SPSS Bootstrapping 19*.

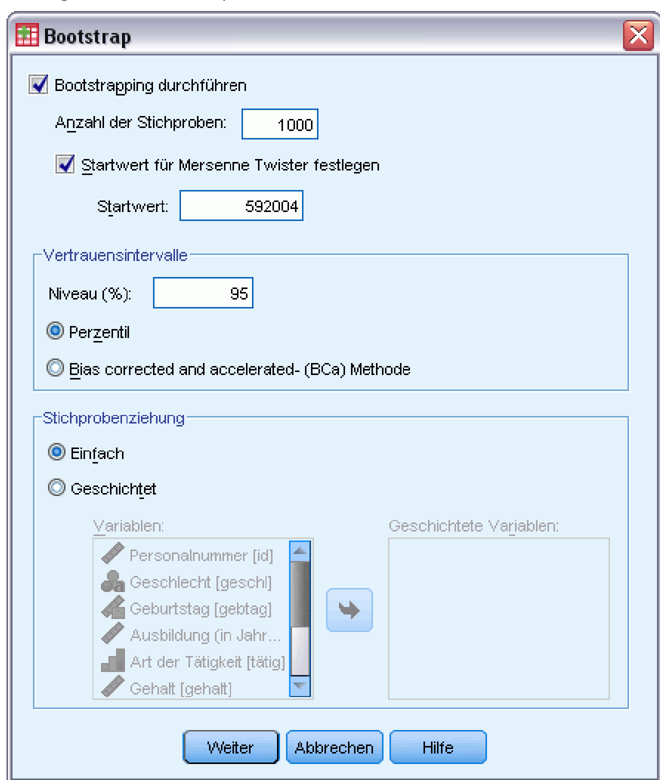
Viele Prozeduren unterstützen das Ziehen von Bootstrap-Stichproben und das Pooling von Ergebnissen aus Analysen von Bootstrap-Stichproben. Steuerelemente für die Angabe von Bootstrap-Analysen sind bei Prozeduren, die Bootstrapping unterstützen, direkt als gemeinsames untergeordnetes Dialogfeld enthalten. Einstellungen im Bootstrap-Dialogfeld werden für sämtliche Prozeduren beibehalten: wenn Sie also über die Dialogfelder eine Häufigkeitenanalyse

mit Bootstrapping durchführen, wird Bootstrapping standardmäßig auch für andere Prozeduren aktiviert, die es unterstützen.

So lassen Sie eine Bootstrap-Analyse berechnen:

▶ Wählen Sie aus den Menüs eine Prozedur aus, die Bootstrapping unterstützt, und klicken Sie auf Bootstrap.

Abbildung 2-1
Dialogfeld "Bootstrap"



▶ Wählen Sie Bootstrapping durchführen.

Optional können Sie folgende Optionen auswählen:

Anzahl der Stichproben. Für das Perzentil und die BCa-Intervalle, die erzeugt werden, empfiehlt es sich, mindestens 1.000 Bootstrap-Stichproben zu verwenden. Geben Sie eine positive Ganzzahl ein.

Startwert für Mersenne-Twister festlegen. Wenn Sie einen Startwert festlegen, können Sie Analysen reproduzieren. Die Verwendung dieses Steuerelements gleicht der Festlegung eines Mersenne-Twisters als aktivem Generator und eines festen Startpunkts für das Dialogfeld "Zufallszahlengeneratoren", mit dem wichtigen Unterschied, dass die Festlegung des Startpunkts in diesem Dialogfeld den aktuellen Status des Zufallszahlengenerators beibehält und diesen Status nach Abschluss der Analyse wiederherstellt.

Konfidenzintervalle. Geben Sie ein Konfidenzniveau größer 50 und kleiner 100 an. Perzentilintervalle verwenden einfach die Bootstrap-Werte, die den gewünschten Konfidenzintervallperzentilen entsprechen. Beispielsweise verwendet ein 95%-Konfidenzintervall die 2,5- und 97,5-Perzentile der Bootstrap-Werte als untere und obere Grenze des Intervalls (bei Bedarf werden die Bootstrap-Werte interpoliert). “Bias corrected and accelerated”- (BCa-) Intervalle sind korrigierte Intervalle, die eine höhere Genauigkeit auf Kosten einer höheren Berechnungszeit bieten.

Stichprobenziehung Die Einfache Methode ist das erneute Ziehen von Fall-Stichproben mit Zurücklegen aus dem ursprünglichen Daten-Set. Die Geschichtete Methode ist das erneute Ziehen von Fall-Stichproben mit Zurücklegen aus dem ursprünglichen Daten-Set *innerhalb* der Schichten, die durch die Kreuzklassifikation von Schichtvariablen definiert werden. Das geschichtete Ziehen von Bootstrap-Stichproben kann von Nutzen sein, wenn die Einheiten innerhalb der Schichten relativ homogen sind, während sich die Einheiten der einzelnen Schichten stark voneinander unterscheiden.

Prozeduren, die Bootstrapping unterstützen

Die folgenden Prozeduren unterstützen Bootstrapping.

Anmerkung:

- Bootstrapping funktioniert nicht bei multiplen imputierten Daten-Sets. Falls es eine Variable *Imputation_* innerhalb des Daten-Sets gibt, wird das Bootstrap-Dialogfeld deaktiviert.
- Bootstrapping verwendet listenweisen Ausschluss, um die Fallbasis zu bestimmen; das bedeutet, dass Fälle mit fehlenden Werten für Analysevariablen von der Analyse ausgeschlossen werden, so dass bei aktivem Bootstrapping auch der listenweise Ausschluss aktiv ist, selbst wenn die Analyseprozedur eine andere Form der Behandlung fehlender Werte vorgibt.

Option “Statistics Base”

Häufigkeiten

- Die Tabelle “Statistik” unterstützt Bootstrap-Schätzer für Mittelwert, Standardabweichung, Varianz, Median, Schiefe, Kurtosis und Perzentile.
- Die Tabelle “Häufigkeiten” unterstützt Bootstrap-Schätzer für Prozent.

Deskriptive Statistik

- Die Tabelle “Deskriptive Statistik” unterstützt Bootstrap-Schätzer für Mittelwert, Standardabweichung, Varianz, Schiefe und Kurtosis.

Explorative Datenanalyse

- Die Tabelle “Deskriptive Statistik” unterstützt Bootstrap-Schätzer für Mittelwert, 5 % getrimmtes Mittel, Standardabweichung, Varianz, Median, Schiefe, Kurtosis und Interquartilbereich.

- Die Tabelle “M-Schätzer” unterstützt Bootstrap-Schätzer für M-Schätzer nach Huber, Tukey-Biweight-Schätzer, M-Schätzer nach Hampel und Andrews-Wellen-Schätzer.
- Die Tabelle “Perzentile” unterstützt Bootstrap-Schätzer für Perzentile.

Kreuztabellen

- Die Tabelle “Richtungsmaße” unterstützt Bootstrap-Schätzer für Lambda, Goodman-und-Kruskal-Tau, Unsicherheitskoeffizient und Somers-d.
- Die Tabelle “Symmetrische Maße” unterstützt Bootstrap-Schätzer für Phi, Cramer-V, Kontingenzkoeffizient, Kendall-Tau-b, Kendall Tau-c, Gamma, Korrelation nach Spearman und Pearson-R.
- Die Tabelle “Risikoschätzer” unterstützt Bootstrap-Schätzer für das Quotenverhältnis.
- Die Tabelle “Gemeinsames Quotenverhältnis nach Mantel-Haenszel” unterstützt Bootstrap-Schätzer und Signifikanztests für ln(Schätzer).

Mittelwerte

- Die Tabelle “Bericht” unterstützt Bootstrap-Schätzer für Mittelwert, Median, Gruppierter Median, Standardabweichung, Varianz, Kurtosis, Schiefe, Harmonisches Mittel und Geometrisches Mittel.

T-Test bei einer Stichprobe

- Die Tabelle “Statistik” unterstützt Bootstrap-Schätzer für Mittelwert und Standardabweichung.
- Die Tabelle “Test” unterstützt Bootstrap-Schätzer und Signifikanztests für die Mittelwertdifferenz.

T-Test bei unabhängigen Stichproben

- Die Tabelle “Gruppenstatistik” unterstützt Bootstrap-Schätzer für Mittelwert und Standardabweichung.
- Die Tabelle “Test” unterstützt Bootstrap-Schätzer und Signifikanztests für die Mittelwertdifferenz.

T-Test bei gepaarten Stichproben

- Die Tabelle “Statistik” unterstützt Bootstrap-Schätzer für Mittelwert und Standardabweichung.
- Die Tabelle “Korrelationen” unterstützt Bootstrap-Schätzer für Korrelationen.
- Die Tabelle “Test” unterstützt Bootstrap-Schätzer für den Mittelwert.

Einfaktorielle ANOVA

- Die Tabelle “Deskriptive Statistik” unterstützt Bootstrap-Schätzer für Mittelwert und Standardabweichung.
- Die Tabelle “Mehrfachvergleiche” unterstützt Bootstrap-Schätzer für die Mittelwertdifferenz.
- Die Tabelle “Kontrasttests” unterstützt Bootstrap-Schätzer und Signifikanztests für den Kontrastwert.

GLM - Univariat

- Die Tabelle “Deskriptive Statistik” unterstützt Bootstrap-Schätzer für Mittelwert und Standardabweichung.
- Die Tabelle “Parameterschätzer” unterstützt Bootstrap-Schätzer und Signifikanztests für den Koeffizienten B.
- Die Tabelle “Kontrastergebnisse” unterstützt Bootstrap-Schätzer und Signifikanztests für die Differenz.
- Die Tabelle “Geschätzte Randmittel: Die Tabelle “Schätzer” unterstützt Bootstrap-Schätzer für den Mittelwert.
- Die Tabelle “Geschätzte Randmittel: Die Tabelle “Paarweise Vergleiche” unterstützt Bootstrap-Schätzer für die Mittelwertdifferenz.
- Die Tabelle “Post-Hoc-Tests: Mehrfachvergleiche” unterstützt Bootstrap-Schätzer für die Mittelwertdifferenz.

Bivariate Korrelationen

- Die Tabelle “Deskriptive Statistik” unterstützt Bootstrap-Schätzer für Mittelwert und Standardabweichung.
- Die Tabelle “Korrelationen” unterstützt Bootstrap-Schätzer für Korrelationen.

Anmerkung: Falls neben Pearson-Korrelationen auch nichtparametrische Korrelationen (Kendall-Tau-b oder Spearman) angefordert werden, fügt das Dialogfeld `CORRELATIONS`- und `NONPAR CORR`-Befehle mit einem separaten `BOOTSTRAP`-Befehl für jeden davon ein. Für die Berechnung aller Korrelationen werden dieselben Bootstrap-Stichproben verwendet.

Partielle Korrelationen

- Die Tabelle “Deskriptive Statistik” unterstützt Bootstrap-Schätzer für Mittelwert und Standardabweichung.
- Die Tabelle “Korrelationen” unterstützt Bootstrap-Schätzer für Korrelationen.

Lineare Regression

- Die Tabelle “Deskriptive Statistik” unterstützt Bootstrap-Schätzer für Mittelwert und Standardabweichung.
- Die Tabelle “Korrelationen” unterstützt Bootstrap-Schätzer für Korrelationen.
- Die Tabelle “Modellzusammenfassung” unterstützt Bootstrap-Schätzer für Durbin-Watson.
- Die Tabelle “Koeffizienten” unterstützt Bootstrap-Schätzer und Signifikanztests für den Koeffizienten B.
- Die Tabelle “Korrelationskoeffizienten” unterstützt Bootstrap-Schätzer für Korrelationen.
- Die Tabelle “Residuenstatistik” unterstützt Bootstrap-Schätzer für Mittelwert und Standardabweichung.

Ordinale Regression

- Die Tabelle “Parameterschätzer” unterstützt Bootstrap-Schätzer und Signifikanztests für den Koeffizienten B.

Diskriminanzanalyse

- Die Tabelle “Standardisierte kanonische Diskriminanzfunktionskoeffizienten” unterstützt Bootstrap-Schätzer für standardisierte Koeffizienten.
- Die Tabelle “Kanonische Diskriminanzfunktionskoeffizienten” unterstützt Bootstrap-Schätzer für nicht standardisierte Koeffizienten.
- Die Tabelle “Klassifizierungsfunktionskoeffizienten” unterstützt Bootstrap-Schätzer für Koeffizienten.

Option “Advanced Statistics”

GLM - Multivariat

- Die Tabelle “Parameterschätzer” unterstützt Bootstrap-Schätzer und Signifikanztests für den Koeffizienten B.

Lineare gemischte Modelle

- Die Tabelle “Schätzungen fester Effekte” unterstützt Bootstrap-Schätzer und Signifikanztests für den Schätzer.
- Die Tabelle “Schätzungen von Kovarianzparametern” unterstützt Bootstrap-Schätzer und Signifikanztests für den Schätzer.

Verallgemeinerte lineare Modelle

- Die Tabelle “Parameterschätzer” unterstützt Bootstrap-Schätzer und Signifikanztests für den Koeffizienten B.

Cox-Regression

- Die Tabelle “Variablen in der Gleichung” unterstützt Bootstrap-Schätzer und Signifikanztests für den Koeffizienten B.

Option “Regression”

Binäre logistische Regression

- Die Tabelle “Variablen in der Gleichung” unterstützt Bootstrap-Schätzer und Signifikanztests für den Koeffizienten B.

Multinomiale logistische Regression

- Die Tabelle “Parameterschätzer” unterstützt Bootstrap-Schätzer und Signifikanztests für den Koeffizienten B.

Zusätzliche Funktionen beim Befehl BOOTSTRAP

Mit der Befehlssyntax können Sie auch Folgendes:

- Ziehen von Residuen- und Wild-Bootstrap-Stichproben (Unterbefehl `SAMPLING`)

Siehe *Befehlssyntaxreferenz* für die vollständigen Syntaxinformationen.

Teil II: Beispiele

Kapitel



Bootstrapping

Bootstrapping ist eine Methode zur Ableitung von robusten Schätzern von Standardfehlern und Konfidenzintervallen für Schätzer wie Mittel, Median, Anteil, Quotenverhältnis, Korrelationskoeffizient oder Regressionskoeffizient. Es kann auch für die Konstruktion von Hypothesentests verwendet werden. Bootstrapping ist besonders als Alternative zu parametrischen Schätzern geeignet, wenn die Annahmen dieser Methoden zweifelhaft (zum Beispiel bei Regressionsmodellen mit heteroskedastischen, auf kleine Stichproben angepassten Residuen) oder parametrische Schlussfolgerungen unmöglich sind oder äußerst komplizierte Formeln zur Berechnung von Standardfehlern erfordern (zum Beispiel bei der Berechnung von Konfidenzintervallen für den Median, Quartilen und andere Perzentilen).

Einsatz von Bootstrapping zum Berechnen von Konfidenzintervallen für Anteile

Eine Telekommunikationsfirma verliert etwa 27 % ihrer Kunden jeden Monat durch Abwanderung. Um bei den Bemühungen zur Verringerung der Abwanderung die richtigen Schwerpunkte setzen zu können, möchte die Geschäftsleitung wissen, ob dieser Prozentsatz zwischen verschiedenen vordefinierten Kundengruppen variiert.

Diese Informationen finden Sie in der Datei *telco.sav*. Für weitere Informationen siehe Thema Beispieldateien in Anhang A auf S. 31. Mit Bootstrapping können Sie ermitteln, ob sich die vier Hauptkundengruppen angemessen mit einer einzigen Abwanderungsquote beschreiben lassen.

Anmerkung: In diesem Beispiel wird die Prozedur "Häufigkeiten" verwendet und die Option "Statistics Base" ist erforderlich.

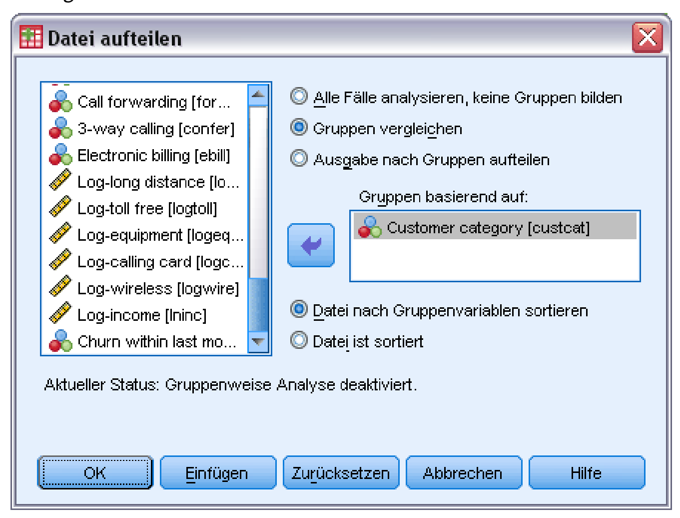
Vorbereitung der Daten

Sie müssen die Datei zunächst nach *Kundenkategorie* aufteilen.

▶ Zur Aufteilung der Datei wählen Sie in den Menüs des Daten-Editors folgende Optionen aus:

Daten > Datei aufteilen...

Abbildung 3-1
Dialogfeld "Datei aufteilen"



▶ Wählen Sie die Option Gruppen vergleichen.

▶ Wählen Sie *Customer category* als die Variable aus, auf der die Gruppen beruhen sollen.

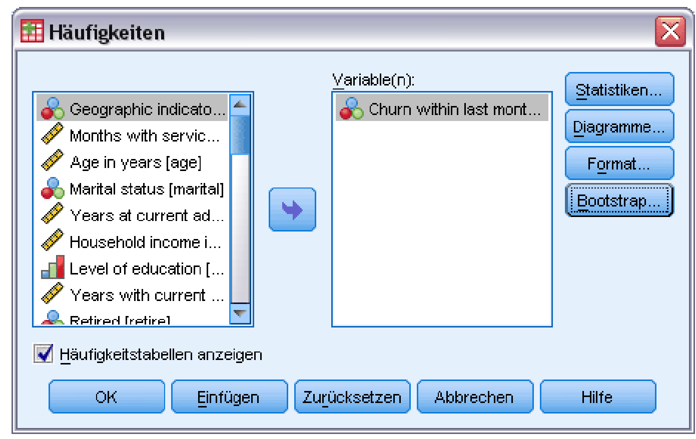
▶ Klicken Sie auf OK.

Durchführen der Analyse

▶ Um Konfidenzintervalle für Anteile zu berechnen, wählen Sie in den Menüs folgende Optionen aus:

Analysieren > Deskriptive Statistiken > Häufigkeiten...

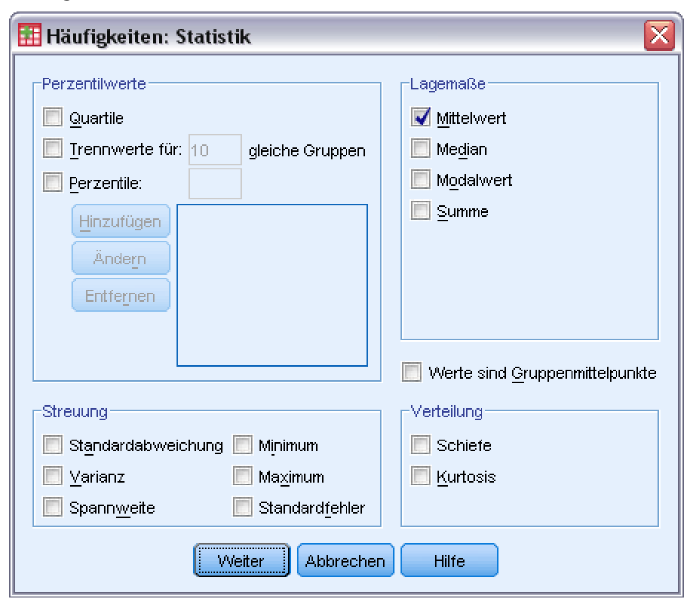
Abbildung 3-2
Dialogfeld "Häufigkeiten"



▶ Wählen Sie *Abwanderung innerhalb des letzten Monats* als Variable in der Analyse aus.

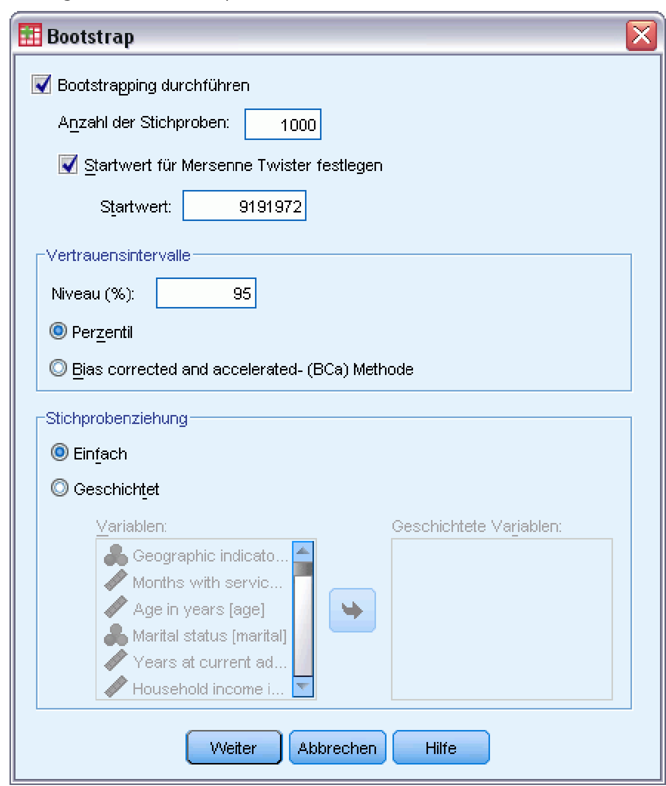
▶ Klicken Sie auf Statistiken.

Abbildung 3-3
Dialogfeld "Statistiken"



- ▶ Wählen Sie Mittelwert in der Gruppe "Lagemaße" aus.
- ▶ Klicken Sie auf Weiter.
- ▶ Klicken Sie im Dialogfeld "Häufigkeiten" auf Bootstrap.

Abbildung 3-4
Dialogfeld "Bootstrap"



- Wählen Sie Bootstrapping durchführen.

- Um die Ergebnisse in diesem Beispiel genau reproduzieren zu können, wählen Sie Startwert für Mersenne Twister festlegen aus und geben Sie 9191972 als Startwert ein.

- Klicken Sie auf Weiter.

- Klicken Sie im Dialogfeld "Häufigkeiten" auf OK.

Diese Auswahl führt zu folgender Befehlssyntax:

```
SORT CASES  BY custcat.
SPLIT FILE LAYERED BY custcat.
PRESERVE.
SET RNG=MT MTINDEX=9191972.
SHOW RNG.
BOOTSTRAP
  /SAMPLING METHOD=SIMPLE
  /VARIABLES INPUT=churn
  /CRITERIA CILEVEL=95 CITYPE=PERCENTILE  NSAMPLES=1000
  /MISSING USERMISSING=EXCLUDE.
FREQUENCIES VARIABLES=churn
  /STATISTICS=MEAN
  /ORDER=ANALYSIS.
RESTORE.
```

- Die Befehle `SORT CASES` und `SPLIT FILE` teilen die Datei an der Variable *custcat* auf.

- Die Befehle PRESERVE und RESTORE “merken” sich den aktuellen Status des Zufallszahlengenerators und stellen das System an dem Punkt nach Ende des Bootstrapping wieder her.
- Der Befehl SET stellt den Zufallszahlengenerator auf den Mersenne Twister und den Index auf den Wert 9191972 ein, damit die Bootstrapping-Ergebnisse exakt reproduziert werden können. Der Befehl SHOW zeigt als Referenz den Index in der Ausgabe an.
- Der Befehl BOOTSTRAP fordert mithilfe einer einfachen Stichprobenziehung 1,000 Bootstrap-Stichproben an.
- Die Variable *Abwanderung* wird verwendet, um die Fallbasis für die Stichprobenziehung zu ermitteln. Datensätze mit fehlenden Werten in dieser Variable werden aus der Analyse entfernt.
- Die Prozedur FREQUENCIES nach BOOTSTRAP wird in jeder Bootstrap-Stichprobe durchgeführt.
- Der Unterbefehl STATISTICS erzeugt in den Originaldaten den Mittelwert für die Variable *churn* (Abwanderung). Zudem werden für den Mittelwert und die Prozentsätze in der Häufigkeitstabelle gepoolte Statistiken erstellt.

Bootstrap-Spezifikationen

Abbildung 3-5
Bootstrap-Spezifikationen

Methode der Stichprobenziehung	Einfach
Anzahl der Stichproben	1000
Konfidenzintervallniveau	95.0%
Konfidenzintervalltyp	Perzentil

Die Tabelle “Boostrap-Spezifikationen” enthält die bei der Stichprobenziehung verwendeten Einstellungen und dient als nützliche Referenz, um zu überprüfen, ob die von Ihnen gewünschte Analyse durchgeführt wurde.

Statistics

Abbildung 3-6
Statistiktabelle mit Bootstrap-Konfidenzintervall für Anteil

Churn within last month

Customer category			Statistic	Bootstrap[a]			
				Verzerrung	Standardfehler	95% Konfidenzintervall	
						Unterer Wert	Oberer Wert
Basic service	N	Gültig	266	0	0	266	266
		Fehlend	0	0	0	0	0
	Mittelwert		.31	.00	.03	.26	.37
E-service	N	Gültig	217	0	0	217	217
		Fehlend	0	0	0	0	0
	Mittelwert		.27	.00	.03	.21	.34
Plus service	N	Gültig	281	0	0	281	281
		Fehlend	0	0	0	0	0
	Mittelwert		.16	.00	.02	.12	.20
Total service	N	Gültig	236	0	0	236	236
		Fehlend	0	0	0	0	0
	Mittelwert		.37	.00	.03	.31	.44

a.Unless otherwise noted, bootstrap results are based on 1000 bootstrap samples

Die Statistiktabelle zeigt für jede Ebene von *Kundenkategorie* den Mittelwert für *Abwanderung innerhalb des letzten Monats* an. Da die *Abwanderung innerhalb des letzten Monats* nur den Wert 0 oder 1 tragen kann, wobei der Wert 1 einem abgewanderten Kunden entspricht, gleicht der Mittelwert dem Anteil der abgewanderten Kunden. Die Spalte Statistik zeigt mithilfe des Original-Daten-Sets die Werte an, die gewöhnlich mit "Häufigkeiten" erzeugt werden. Die Spalten Bootstrap werden anhand der Bootstrapping-Algorithmen erzeugt.

- Verzerrung ist die Differenz zwischen dem Durchschnittswert dieser Statistik in allen Bootstrap-Stichproben und dem Wert in der Spalte Statistik. In diesem Fall wird der Mittelwert von *Abwanderung innerhalb des letzten Monats* für alle 1000 Bootstrap-Stichproben berechnet, und anschließend wird der Durchschnitt dieser Mittelwerte berechnet.
- "Std. "Fehler" ist der Standardfehler des Mittelwerts von *Abwanderung innerhalb des letzten Monats* in allen 1000 Bootstrap-Stichproben.
- Die Untergrenze des 95 %-Bootstrap-Konfidenzintervalls ist eine Interpolation des 25. und 26. Mittelwerts von *Abwanderung innerhalb des letzten Monats,* wenn die 1000 Bootstrap-Stichproben in aufsteigender Reihenfolge sortiert werden. Die Obergrenze ist eine Interpolation des 975. und 976. Mittelwerts.

Die Ergebnisse in der Tabelle lassen darauf schließen, dass die Abwanderungsquote je nach Kundengruppe unterschiedlich ist. Insbesondere die Tatsache, dass sich das Konfidenzintervall für *Plus service*-Kunden nicht mit anderen überschneidet, lässt darauf schließen, dass diese Kunden im Durchschnitt mit einer geringeren Wahrscheinlichkeit abwandern.

Bei der Arbeit mit kategorialen Variablen mit nur zwei Werten bieten diese Konfidenzintervalle eine Alternative zu jenen, die mit der Prozedur "Nichtparametrische Tests bei einer Stichprobe" oder "T-Tests bei einer Stichprobe" erzeugt werden.

Häufigkeitstabelle (Correspondence Analysis)

Abbildung 3-7
Häufigkeitstabelle mit Bootstrap-Konfidenzintervall für Anteil

Customer category			Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente	Bootstrap für Prozent[a]			
							Verzerrung	Standardfehler	95% Konfidenzintervall	
									Unterer Wert	Oberer Wert
Basic service	Gültig	No	183	68.8	68.8	68.8	.0	2.8	63.2	74.4
		Yes	83	31.2	31.2	100.0	.0	2.8	25.6	36.8
		Gesamt	266	100.0	100.0		.0	.0	100.0	100.0
E-service	Gültig	No	158	72.8	72.8	72.8	.1	3.1	66.4	78.8
		Yes	59	27.2	27.2	100.0	-.1	3.1	21.2	33.6
		Gesamt	217	100.0	100.0		.0	.0	100.0	100.0
Plus service	Gültig	No	237	84.3	84.3	84.3	.0	2.1	80.1	88.3
		Yes	44	15.7	15.7	100.0	.0	2.1	11.7	19.9
		Gesamt	281	100.0	100.0		.0	.0	100.0	100.0
Total service	Gültig	No	148	62.7	62.7	62.7	.0	3.2	56.4	69.1
		Yes	88	37.3	37.3	100.0	.0	3.2	30.9	43.6
		Gesamt	236	100.0	100.0		.0	.0	100.0	100.0

a.Unless otherwise noted, bootstrap results are based on 1000 bootstrap samples

Die Häufigkeitstabelle zeigt Konfidenzintervalle für die Prozentsätze (Anteil × 100 %) für jede Kategorie und diese sind daher für alle kategorialen Variablen verfügbar. Vergleichbare Konfidenzintervalle sind nicht an anderer Stelle in diesem Produkt verfügbar.

Einsatz von Bootstrapping zum Berechnen von Konfidenzintervallen für Mediane

Bei der Durchsicht von Mitarbeiterdaten ist die Geschäftsleitung an der bisherigen Arbeitserfahrung seiner Mitarbeiter interessiert. Die Arbeitserfahrung ist rechtslastig, was bedeutet, dass der Mittelwert eine weniger wünschenswerte Schätzung der "typischen" bisherigen Arbeitserfahrung unter Mitarbeitern darstellt als der Median. Ohne Bootstrapping sind Konfidenzintervalle für den Median jedoch in den statistischen Prozeduren im Produkt nicht allgemein verfügbar.

Diese Informationen finden Sie in der Datei *Employee data.sav*. Für weitere Informationen siehe Thema Beispieldateien in Anhang A auf S. 31. Einsatz von Bootstrapping zum Berechnen von Konfidenzintervallen für den Median.

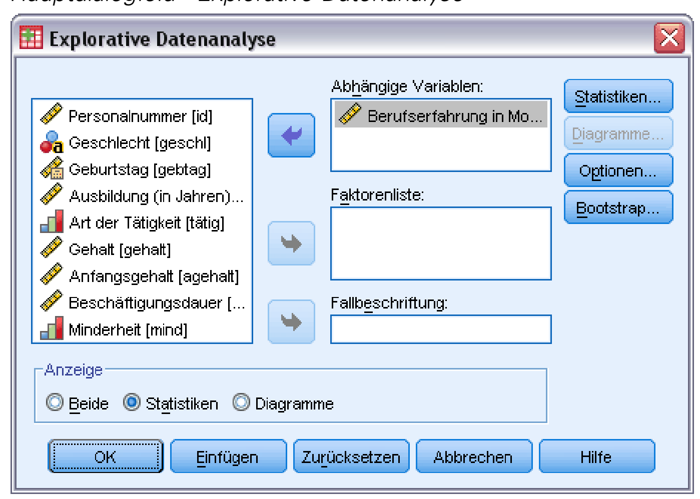
Anmerkung: In diesem Beispiel wird die Prozedur "Explorative Datenanalyse" verwendet und die Option "Statistics Base" ist erforderlich.

Durchführen der Analyse

▶ Um Konfidenzintervalle für den Median zu berechnen, wählen Sie in den Menüs folgende Optionen aus:

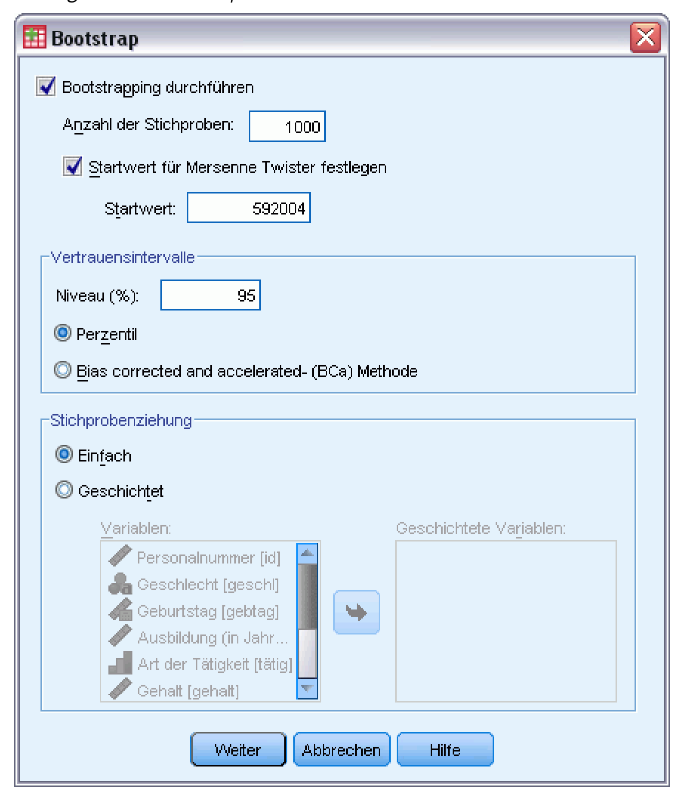
Analysieren > Deskriptive Statistiken > Explorative Datenanalyse...

Abbildung 3-8
Hauptdialogfeld "Explorative Datenanalyse"



- Wählen Sie *Bisherige Erfahrung (Monate) [bisherf]* als abhängige Variable aus.
- Wählen Sie Statistik in der Gruppe "Anzeige" aus.
- Klicken Sie auf Bootstrap.

Abbildung 3-9
Dialogfeld "Bootstrap"



- Wählen Sie Bootstrapping durchführen.
- Um die Ergebnisse in diesem Beispiel genau reproduzieren zu können, wählen Sie Startwert für Mersenne Twister festlegen aus und geben Sie 592004 als Startwert ein.
- Um genauere Intervalle (auf Kosten zusätzlicher Verarbeitungszeit) zu berechnen, wählen Sie Bias corrected and accelerated (BCa)-Methode aus.
- Klicken Sie auf Weiter.
- Klicken Sie im Dialogfeld "Explorative Datenanalyse" auf OK.

Diese Auswahl führt zu folgender Befehlssyntax:

```
PRESERVE.
SET RNG=MT MTINDEX=592004.
SHOW RNG.
BOOTSTRAP
  /SAMPLING METHOD=SIMPLE
  /VARIABLES TARGET=prevexp
  /CRITERIA CILEVEL=95 CITYPE=BCA  NSAMPLES=1000
  /MISSING USERMISSING=EXCLUDE.
EXAMINE VARIABLES=prevexp
  /PLOT NONE
  /STATISTICS DESCRIPTIVES
  /CINTERVAL 95
  /MISSING LISTWISE
  /NOTOTAL.
```

```
RESTORE.
```

- Die Befehle `PRESERVE` und `RESTORE` “merken” sich den aktuellen Status des Zufallszahlengenerators und stellen das System an dem Punkt nach Ende des Bootstrapping wieder her.
- Der Befehl `SET` stellt den Zufallszahlengenerator auf den Mersenne Twister und den Index auf den Wert 592004 ein, damit die Bootstrapping-Ergebnisse exakt reproduziert werden können. Der Befehl `SHOW` zeigt als Referenz den Index in der Ausgabe an.
- Der Befehl `BOOTSTRAP` fordert mithilfe einer einfachen Stichprobenziehung 1000 Bootstrap-Stichproben an.
- Der Unterbefehl `VARIABLES` legt fest, dass die Variable *bisherf* zur Ermittlung der Fallbasis für die Stichprobenziehung verwendet wird. Datensätze mit fehlenden Werten in dieser Variable werden aus der Analyse entfernt.
- Der Unterbefehl `CRITERIA` fordert zusätzlich zur Anzahl der Bootstrap-Stichproben “Bias-corrected and accelerated”-Bootstrap-Konfidenzintervalle anstelle der standardmäßigen Perzentilintervalle an.
- Die Prozedur `EXAMINE` nach `BOOTSTRAP` wird in jeder Bootstrap-Stichprobe durchgeführt.
- Der Unterbefehl `PLOT` deaktiviert die Diagrammausgabe.
- Für alle anderen Optionen gelten die Standardwerte.

Deskriptive Statistik

Abbildung 3-10
Tabelle “Deskriptive Statistik” mit Bootstrap-Konfidenzintervallen

					Bootstrap[a]			
							BCa 95% Konfidenzintervall	
			Statistik	Standardfehler	Verzerrung	Standardfehler	Unterer Wert	Oberer Wert
Berufserfahrung in Monaten	Mittelwert		95.86	4.804	-.01	4.86	86.39	105.20
	95% Konfidenzintervall des Mittelwerts	Untergrenze	86.42					
		Obergrenze	105.30					
	5% getrimmtes Mittel		84.64		.02	4.94	75.38	94.21
	Median		55.00		-.11	3.66	50.00	60.00
	Varianz		10938.281		18.783	977.081	8954.509	13057.229
	Standardabweichung		104.586		-.015	4.689	94.644	114.245
	Minimum		0					
	Maximum		476					
	Spannweite		476					
	Interquartilbereich		121		-1	10	103	137
	Schiefe		1.510	.112	.006	.110	1.284	1.768
	Kurtosis		1.696	.224	.040	.463	.823	2.876

a.Unless otherwise noted, bootstrap results are based on 1000 bootstrap samples

Die Tabelle “Deskriptive Statistik” enthält zahlreiche Statistiken und Bootstrap-Intervalle für diese Statistiken. Das Bootstrap-Konfidenzintervall für den Mittelwert (86,39; 105,20) gleicht dem parametrischen Konfidenzintervall (86,42; 105,30), was darauf schließen lässt, dass der “typische” Mitarbeiter über rund 7-9 Jahre an bisheriger Erfahrung verfügt. Die Variable *Bisherige Erfahrung (Monate)* weist jedoch eine schiefe Verteilung auf, was bedeutet, dass der Mittelwert einen weniger wünschenswerten Indikator eines “typischen” derzeitigen Gehalts darstellt als der Median. Das Bootstrap-Konfidenzintervall für den Median (50,00; 60,00) ist schmaler und hat

einen niedrigeren Wert als das Konfidenzintervall für den Mittelwert, was darauf schließen lässt, dass der "typische" Mitarbeiter über rund 4-5 Jahre an bisheriger Erfahrung verfügt. Durch den Einsatz von Bootstrapping können Werte berechnet werden, die die typische bisherige Erfahrung besser darstellen.

Einsatz von Boostrapping zur Auswahl besserer Einflussvariablen

Bei der Durchsicht von Mitarbeiterdaten ist die Geschäftsleitung daran interessiert zu ermitteln, welche Faktoren Gehaltserhöhungen von Mitarbeitern entsprechen, indem ein lineares Modell über die Differenz zwischen aktuellem und Anfangsgehalt erstellt wird. Beim Bootstrapping eines linearen Modells können Sie spezielle Resampling-Methoden (Residuen- und Wild-Bootstrap) verwenden, um genauere Ergebnisse zu erzielen.

Diese Informationen finden Sie in der Datei *Employee data.sav*. Für weitere Informationen siehe Thema Beispieldateien in Anhang A auf S. 31.

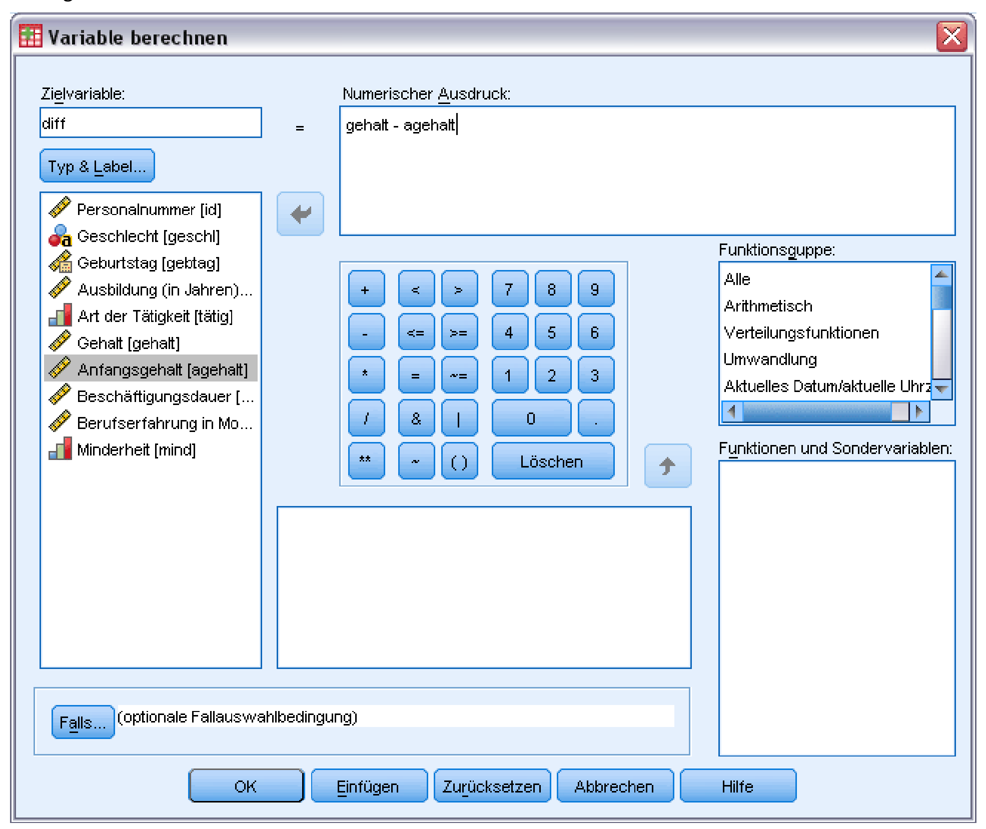
Anmerkung: In diesem Beispiel wird die Prozedur "GLM - Univariat" verwendet und die Option "Statistics Base" ist erforderlich.

Vorbereitung der Daten

Sie müssen zunächst die Differenz zwischen aktuellem und Anfangsgehalt berechnen.

▶ Wählen Sie die folgenden Befehle aus den Menüs aus:
Transformieren > Variable berechnen...

Abbildung 3-11
Dialogfeld "Variable berechnen"



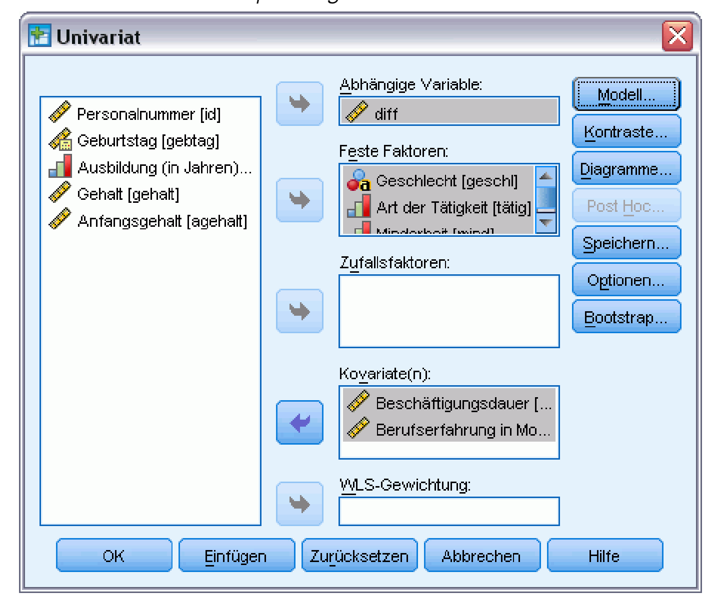
- Geben Sie diff als Zielvariable ein.
- Geben Sie Gehalt-Anf.gehalt als numerischen Ausdruck ein.
- Klicken Sie auf OK.

Durchführen der Analyse

Um "GLM - Univariat" mit Wild-Residuum-Bootstrapping auszuführen, müssen Sie zuerst Residuen erstellen.

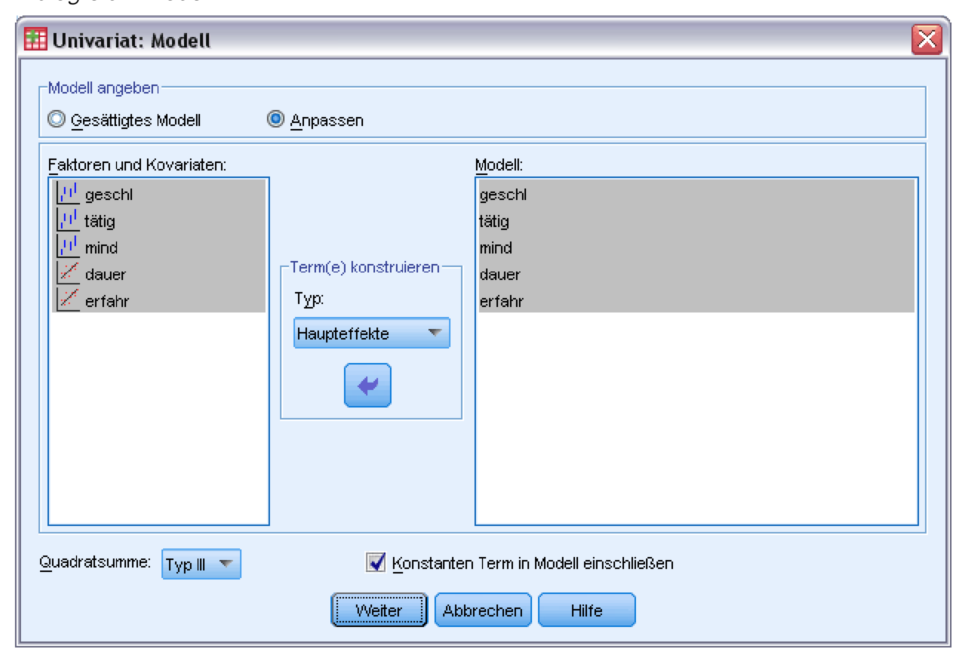
- Wählen Sie die folgenden Befehle aus den Menüs aus:
  Analysieren > Allgemeines lineares Modell > Univariat...

Abbildung 3-12
GLM - Univariat – Hauptdialogfeld



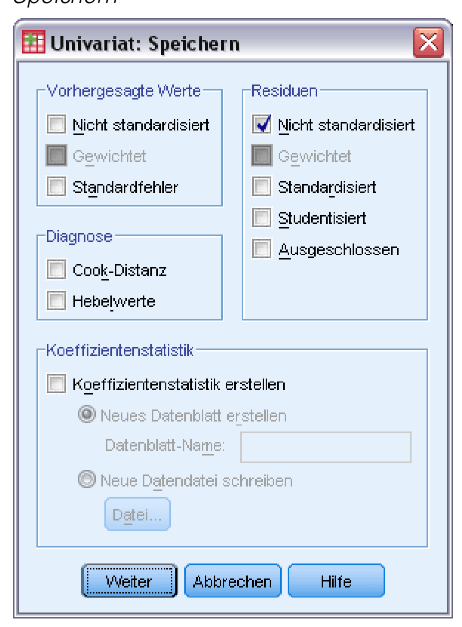
- Wählen Sie *diff* als abhängige Variable aus.
- Wählen Sie *Geschlecht [geschl]*, *Art der Tätigkeit [tätig]* und *Minderheit [minderh]* als feste Faktoren aus.
- Wählen Sie *Beschäftigungsdauer [dauer]* und *Bisherige Erfahrung (Monate) [bisherf]* als Kovariaten aus.
- Klicken Sie auf Modell.

Abbildung 3-13
Dialogfeld "Modell"



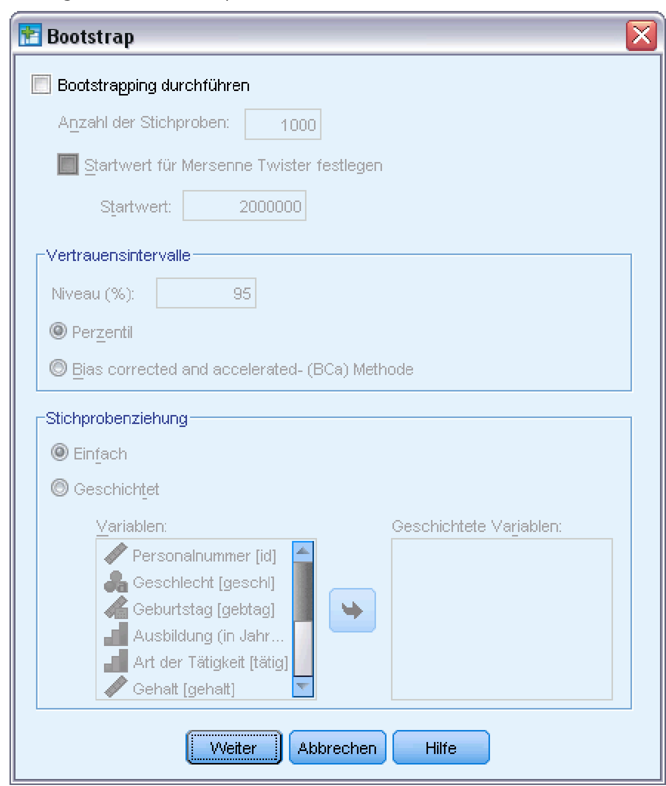
- Wählen Sie Benutzerdefiniert und anschließend Haupteffekte aus der Dropdown-Liste "Term(e) konstruieren" aus.
- Wählen Sie *geschl* bis *bisherf* als Modellterme aus.
- Klicken Sie auf Weiter.
- Klicken Sie im Dialogfeld "GLM - Univariat" auf Speichern.

Abbildung 3-14
Speichern"



- Wählen Sie in der Gruppe "Residuen" die Option Nicht standardisiert.
- Klicken Sie auf Weiter.
- Klicken Sie im Dialogfeld "GLM - Univariat" auf Bootstrap.

Abbildung 3-15
Dialogfeld "Bootstrap"



Die Bootstrap-Einstellungen sind in allen Dialogfeldern gültig, die Bootstrapping unterstützen. Das Speichern neuer Variablen in das Daten-Set wird bei aktiviertem Bootstrapping nicht unterstützt, stellen Sie daher sicher, dass es deaktiviert ist.

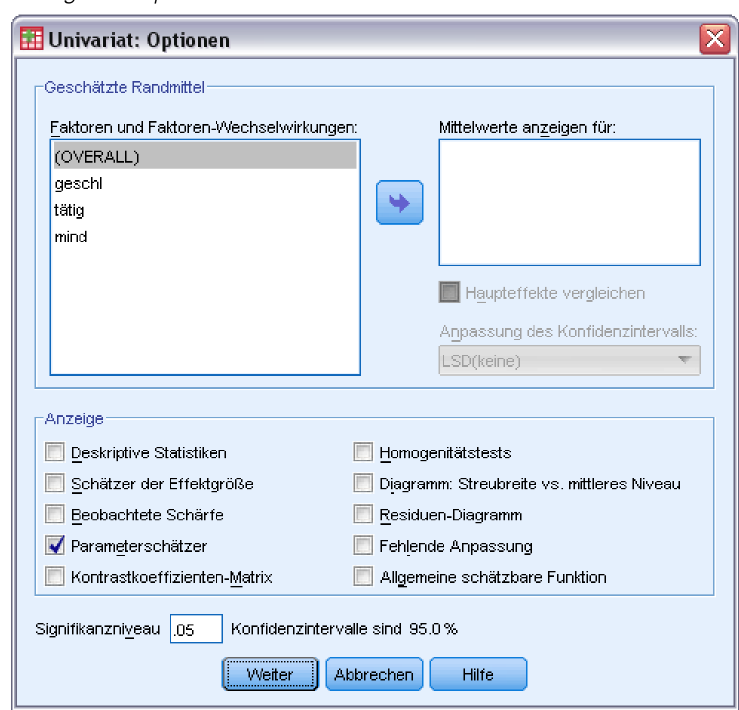
▶ Deaktivieren Sie falls nötig Bootstrapping durchführen.

▶ Klicken Sie im Dialogfeld "GLM - Univariat" auf OK. Das Daten-Set enthält nun eine neue Variable, *RES_1*, die die nicht standardisierten Residuen aus diesem Modell enthält.

▶ Rufen Sie das Dialogfeld "GLM - Univariat" erneut auf und klicken Sie auf Speichern.

▶ Deaktivieren Sie Nicht standardisiert, klicken Sie auf Weiter und anschließend auf Optionen im Dialogfeld “GLM - Univariat”.

Abbildung 3-16
Dialogfeld “Optionen”

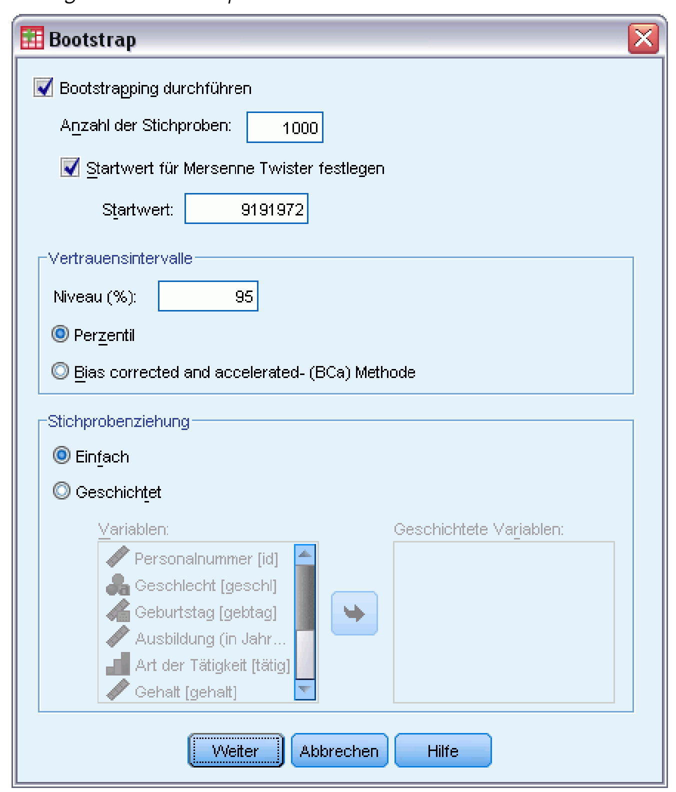


▶ Wählen Sie in der Gruppe “Anzeige” Parameterschätzer aus.

▶ Klicken Sie auf Weiter.

▶ Klicken Sie im Dialogfeld “GLM - Univariat” auf Bootstrap.

Abbildung 3-17
Dialogfeld "Bootstrap"



- Wählen Sie Bootstrapping durchführen.
- Um die Ergebnisse in diesem Beispiel genau reproduzieren zu können, wählen Sie Startwert für Mersenne Twister festlegen aus und geben Sie 9191972 als Startwert ein.
- Da keine Optionen zum Durchführen von Wild-Bootstrapping in den Dialogfeldern vorhanden sind, klicken Sie im Dialogfeld "GLM - Univariat" auf Weiter und anschließend auf Einfügen.

  Diese Auswahl führt zu folgender Befehlssyntax:

```
PRESERVE.
SET RNG=MT MTINDEX=9191972.
SHOW RNG.
BOOTSTRAP
 /SAMPLING METHOD=SIMPLE
 /VARIABLES TARGET=diff INPUT=gender jobcat minority jobtime prevexp
 /CRITERIA CILEVEL=95 CITYPE=PERCENTILE  NSAMPLES=1000
 /MISSING USERMISSING=EXCLUDE.
UNIANOVA diff BY gender jobcat minority WITH jobtime prevexp
 /METHOD=SSTYPE(3)
 /INTERCEPT=INCLUDE
 /PRINT=PARAMETER
 /CRITERIA=ALPHA(.05)
 /DESIGN=gender jobcat minority jobtime prevexp.
```

```
RESTORE.
```

Um das Ziehen von Wild-Bootstrap-Stichproben durchzuführen, bearbeiten Sie das Schlüsselwort `METHOD` im Unterbefehl `SAMPLING` so, dass `METHOD=WILD(RESIDUALS=RES_1)` zu lesen ist.

Die “endgültige” Befehlssyntaxreihe sieht dann wie folgt aus:

```
PRESERVE.
SET RNG=MT MTINDEX=9191972.
SHOW RNG.
BOOTSTRAP
 /SAMPLING METHOD=WILD(RESIDUALS=RES_1)
 /VARIABLES TARGET=diff INPUT=gender jobcat minority jobtime prevexp
 /CRITERIA CILEVEL=95 CITYPE=PERCENTILE  NSAMPLES=1000
 /MISSING USERMISSING=EXCLUDE.
UNIANOVA diff BY gender jobcat minority WITH jobtime prevexp
 /METHOD=SSTYPE(3)
 /INTERCEPT=INCLUDE
 /PRINT=PARAMETER
 /CRITERIA=ALPHA(.05)
 /DESIGN=gender jobcat minority jobtime prevexp.
RESTORE.
```

- Die Befehle `PRESERVE` und `RESTORE` “merken” sich den aktuellen Status des Zufallszahlengenerators und stellen das System an dem Punkt nach Ende des Bootstrapping wieder her.
- Der Befehl `SET` stellt den Zufallszahlengenerator auf den Mersenne Twister und den Index auf den Wert 9191972 ein, damit die Bootstrapping-Ergebnisse exakt reproduziert werden können. Der Befehl `SHOW` zeigt als Referenz den Index in der Ausgabe an.
- Der Befehl `BOOTSTRAP` fordert mithilfe einer Wild-Stichprobenziehung und *RES_1* als Variable mit den Residuen 1000 Bootstrap-Stichproben an.
- Der Unterbefehl `VARIABLES` legt *diff* als Zielvariable in dem linearen Modell fest; sie wird zusammen mit den Variablen *geschl*, *tätig*, *minderh*, *dauer* und *bisherf* zur Ermittlung der Fallbasis für die Stichprobenziehung verwendet. Datensätze mit fehlenden Werten in diesen Variablen werden aus der Analyse entfernt.
- Der Unterbefehl `CRITERIA` fordert zusätzlich zur Anzahl der Bootstrap-Stichproben “Bias-corrected and accelerated”-Bootstrap-Konfidenzintervalle anstelle der standardmäßigen Perzentilintervalle an.
- Die Prozedur `UNIANOVA` nach `BOOTSTRAP` wird bei jeder Bootstrap-Stichprobe durchgeführt und erzeugt Parameterschätzer für die Originaldaten. Zudem werden für die Modellkoeffizienten gepoolte Statistiken erstellt.

Parameter-Schätzer

Abbildung 3-18
Parameterschätzer

Abhängige Variable:diff

Parameter	Regressionskoeffizient B	Standardfehler	T	Sig.	95%-Konfidenzintervall	
					Untergrenze	Obergrenze
Konstanter Term	18703.761	2961.969	6.315	.000	12883.323	24524.199
[geschl=m]	4085.253	726.416	5.624	.000	2657.804	5512.701
[geschl=w]	0[a]	.	.	.	.	.
[tätig=1]	-17717.706	939.798	-18.853	.000	-19564.463	-15870.949
[tätig=2]	-13101.918	1780.683	-7.358	.000	-16601.061	-9602.776
[tätig=3]	0[a]	.	.	.	.	.
[mind=0]	1332.363	819.349	1.626	.105	-277.705	2942.431
[mind=1]	0[a]	.	.	.	.	.
dauer	145.539	32.586	4.466	.000	81.505	209.572
erfahr	-21.423	3.575	-5.993	.000	-28.447	-14.398

a.Dieser Parameter wird auf Null gesetzt, weil er redundant ist.

Die Tabelle “Parameterschätzer” zeigt die gewöhnlichen Nicht-Bootstrap-Parameterschätzer für die Modellterme an. Der Signifikanzwert 0,105 für *[minderh=0]* ist größer als 0,05, was darauf schließen lässt, dass *Minderheit* keine Auswirkungen auf Gehaltserhöhungen hat.

Abbildung 3-19
Bootstrap-Parameterschätzer

Abhängige Variable:diff

Parameter	Regressionskoeffizient B	Bootstrap[a]				
		Verzerrung	Standardfehler	Sig. (2-seitig)	95%-Konfidenzintervall	
					Unterer Wert	Oberer Wert
Konstanter Term	18703.761	-62.604	3330.877	.001	12141.023	24980.359
[geschl=m]	4085.253	-32.480	622.971	.001	2892.131	5365.321
[geschl=w]	0	0	0		0	0
[tätig=1]	-17717.706	46.324	1454.230	.001	-20671.451	-14889.507
[tätig=2]	-13101.918	47.958	1753.311	.001	-16658.596	-9671.891
[tätig=3]	0	0	0		0	0
[mind=0]	1332.363	-10.592	651.144	.012	57.831	2642.534
[mind=1]	0	0	0		0	0
dauer	145.539	.707	35.285	.001	79.081	217.761
erfahr	-21.423	-.065	2.859	.001	-27.533	-16.055

a.Unless otherwise noted, bootstrap results are based on 1000 bootstrap samples

Sehen Sie sich nun die Tabelle “Bootstrap für Parameterschätzer” an. In der Spalte “Std. fehler” kann man erkennen, dass die parametrischen Standardfehler bei manchen Koeffizienten wie dem konstanten Term im Vergleich zu den Bootstrap-Schätzern zu klein sind und daher die Konfidenzintervalle größer sind. Bei manchen Koeffizienten wie *[minderh=0]* waren die parametrischen Standardfehler zu groß, während der Signifikanzwert 0.006, der in den Bootstrap-Ergebnissen angezeigt wurde, unter 0,05 lag, was zeigt, dass die beobachtete Differenz bei Gehaltserhöhungen zwischen Mitarbeitern, die Minderheiten darstellen oder nicht, nicht auf Zufall zurückzuführen ist. Die Geschäftsleitung weiß nun, dass diese Differenz genauer analysiert werden muss, um mögliche Ursachen zu finden.

Empfohlene Literatur

In den folgenden Texten finden Sie weitere Informationen über Bootstrapping:

Davison, A. C., als auch D. V. Hinkley. 2006. *Bootstrap Methods and their Application*. : Cambridge University Press.

Shao, J., als auch D. Tu. 1995. *The Jackknife and Bootstrap*. New York: Springer.

Anhang



Beispieldateien

Die zusammen mit dem Produkt installierten Beispieldateien finden Sie im Unterverzeichnis *Samples* des Installationsverzeichnisses. Für jeder der folgenden Sprachen gibt es einen eigenen Ordner innerhalb des Unterverzeichnisses "Samples": Englisch, Französisch, Deutsch, Italienisch, Japanisch, Koreanisch, Polnisch, Russisch, Vereinfachtes Chinesisch, Spanisch und Traditionelles Chinesisch.

Nicht alle Beispieldateien stehen in allen Sprachen zur Verfügung. Wenn eine Beispieldatei nicht in einer Sprache zur Verfügung steht, enthält der jeweilige Sprachordner eine englische Version der Beispieldatei.

Beschreibungen

Im Folgenden finden Sie Kurzbeschreibungen der in den verschiedenen Beispielen in der Dokumentation verwendeten Beispieldateien.

- **accidents.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, bei der es um eine Versicherungsgesellschaft geht, die alters- und geschlechtsabhängige Risikofaktoren für Autounfälle in einer bestimmten Region untersucht. Jeder Fall entspricht einer Kreuzklassifikation von Alterskategorie und Geschlecht.
- **adl.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, bei der es um Bemühungen geht, die Vorteile einer vorgeschlagenen Therapieform für Schlaganfallpatienten zu ermitteln. Ärzte teilten weibliche Schlaganfallpatienten nach dem Zufallsprinzip jeweils einer von zwei Gruppen zu. Die erste Gruppe erhielt die physische Standardtherapie, die zweite erhielt eine zusätzliche Emotionaltherapie. Drei Monate nach den Behandlungen wurden die Fähigkeiten der einzelnen Patienten, übliche Alltagsaktivitäten auszuführen, als ordinale Variablen bewertet.
- **advert.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, bei der es um die Bemühungen eines Einzelhändlers geht, die Beziehungen zwischen den in Werbung investierten Beträgen und den daraus resultierenden Umsätzen zu untersuchen. Zu diesem Zweck hat er die Umsätze vergangener Jahre und die zugehörigen Werbeausgaben zusammengestellt.
- **aflatoxin.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, bei der es um Tests von Maisernten auf Aflatoxin geht, ein Gift, dessen Konzentration stark zwischen und innerhalb von Ernteerträgen schwankt. Ein Kornverarbeitungsbetrieb hat aus 8 Ernteerträgen je 16 Proben erhalten und das Aflatoxinniveau in Teilen pro Milliarde (parts per billion, PPB) gemessen.
- **aflatoxin20.sav.** Diese Datendatei enthält die Aflatoxinmessungen aus jeder der 16 Stichproben aus den Erträgen 4 und 8 der Datendatei *aflatoxin.sav*.

- **anorectic.sav.** Bei der Ausarbeitung einer standardisierten Symptomatologie anorektischen/bulimischen Verhaltens führten Forscher (Van der Ham, Meulman, Van Strien, als auch Van Engeland, 1997)) eine Studie mit 55 Jugendlichen mit bekannten Ess-Störungen durch. Jeder Patient wurde vier Mal über einen Zeitraum von vier Jahren untersucht, es fanden also insgesamt 220 Beobachtungen statt. Bei jeder Beobachtung erhielten die Patienten Scores für jedes von 16 Symptomen. Die Symptomwerte fehlen für Patient 71 zum Zeitpunkt 2, Patient 76 zum Zeitpunkt 2 und Patient 47 zum Zeitpunkt 3, wodurch 217 gültige Beobachtungen verbleiben.
- **autoaccidents.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, bei der es um die Bemühungen eines Versicherungsanalysten geht, ein Modell zur Anzahl der Autounfälle pro Fahrer unter Berücksichtigung von Alter und Geschlecht zu erstellen. Jeder Fall stellt einen Fahrer dar und erfasst das Geschlecht des Fahrers, sein Alter in Jahren und die Anzahl der Autounfälle in den letzten fünf Jahren.
- **band.sav.** Diese Datendatei enthält die hypothetischen wöchentlichen Verkaufszahlen von CDs für eine Musikgruppe. Daten für drei mögliche Einflussvariablen wurden ebenfalls aufgenommen.
- **bankloan.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, bei der es um die Bemühungen einer Bank geht, den Anteil der nicht zurückgezahlten Kredite zu reduzieren. Die Datei enthält Informationen zum Finanzstatus und demografischen Hintergrund von 850 früheren und potenziellen Kunden. Bei den ersten 700 Fällen handelt es sich um Kunden, denen bereits ein Kredit gewährt wurde. Bei den letzten 150 Fällen handelt es sich um potenzielle Kunden, deren Kreditrisiko die Bank als gering oder hoch einstufen möchte.
- **bankloan_binning.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, die Informationen zum Finanzstatus und demografischen Hintergrund von 5.000 früheren Kunden enthält.
- **behavior.sav.** In einem klassischen Beispiel ((Price als auch Bouffard, 1974)) wurden 52 Schüler/Studenten gebeten, die Kombinationen aus 15 Situationen und 15 Verhaltensweisen auf einer 10-Punkte-Skala von 0 = "ausgesprochen angemessen" bis 9 = "ausgesprochen unangemessen" zu bewerten. Die Werte werden über die einzelnen Personen gemittelt und als Unähnlichkeiten verwendet.
- **behavior_ini.sav.** Diese Datendatei enthält eine Ausgangskonfiguration für eine zweidimensionale Lösung für *behavior.sav*.
- **brakes.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, bei der es um die Qualitätskontrolle in einer Fabrik geht, die Scheibenbremsen für Hochleistungsautomobile herstellt. Die Datendatei enthält Messungen des Durchmessers von 16 Scheiben aus 8 Produktionsmaschinen. Der Zieldurchmesser für die Scheiben ist 322 Millimeter.
- **breakfast.sav.** In einer klassischen Studie ((Green als auch Rao, 1972)) wurden 21 MBA-Studenten der Wharton School mit ihren Lebensgefährten darum gebeten, 15 Frühstücksartikel in der Vorzugsreihenfolge von 1 = "am meisten bevorzugt" bis 15 = "am wenigsten bevorzugt" zu ordnen. Die Bevorzugungen wurden in sechs unterschiedlichen Szenarien erfasst, von "Overall preference" (Allgemein bevorzugt) bis "Snack, with beverage only" (Imbiss, nur mit Getränk).
- **breakfast-overall.sav.** Diese Datei enthält die Daten zu den bevorzugten Frühstücksartikeln, allerdings nur für das erste Szenario, "Overall preference" (Allgemein bevorzugt).

- **broadband_1.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, die die Anzahl der Abonnenten eines Breitband-Service, nach Region geordnet, enthält. Die Datendatei enthält die monatlichen Abonnentenzahlen für 85 Regionen über einen Zeitraum von vier Jahren.
- **broadband_2.sav** Diese Datendatei stimmt mit *broadband_1.sav* überein, enthält jedoch Daten für weitere drei Monate.
- **car_insurance_claims.sav.** Ein an anderer Stelle ((McCullagh als auch Nelder, 1989)) vorgestelltes und analysiertes Daten-Set bezieht sich auf Schadensansprüche für Autos. Die durchschnittliche Höhe der Schadensansprüche lässt sich mit Gamma-Verteilung modellieren. Dazu wird eine inverse Verknüpfungsfunktion verwendet, um den Mittelwert der abhängigen Variablen mit einer linearen Kombination aus Alter des Versicherungsnehmers, Fahrzeugtyp und Fahrzeugalter in Bezug zu setzen. Die Anzahl der eingereichten Schadensansprüche kann als Skalierungsgewicht verwendet werden.
- **car_sales.sav.** Diese Datendatei enthält hypothetische Verkaufsschätzer, Listenpreise und physische Spezifikationen für verschiedene Fahrzeugfabrikate und -modelle. Die Listenpreise und physischen Spezifikationen wurden von *edmunds.com* und Hersteller-Websites entnommen.
- **car_sales_uprepared.sav.** Hierbei handelt es sich um eine modifizierte Version der Datei *car_sales.sav*, die keinerlei transformierte Versionen der Felder enthält.
- **carpet.sav** In einem beliebten Beispiel möchte (Green als auch Wind, 1973) einen neuen Teppichreiniger vermarkten und dazu den Einfluss von fünf Faktoren auf die Bevorzugung durch den Verbraucher untersuchen: Verpackungsgestaltung, Markenname, Preis, Gütesiegel, *Good Housekeeping* und Geld-zurück-Garantie. Die Verpackungsgestaltung setzt sich aus drei Faktorenebenen zusammen, die sich durch die Position der Auftragebürste unterscheiden. Außerdem gibt es drei Markennamen (*K2R*, *Glory* und *Bissell*), drei Preisstufen sowie je zwei Ebenen (Nein oder Ja) für die letzten beiden Faktoren. 10 Kunden stufen 22 Profile ein, die durch diese Faktoren definiert sind. Die Variable *Preference* enthält den Rang der durchschnittlichen Einstufung für die verschiedenen Profile. Ein niedriger Rang bedeutet eine starke Bevorzugung. Diese Variable gibt ein Gesamtmaß der Bevorzugung für die Profile an.
- **carpet_prefs.sav.** Diese Datendatei beruht auf denselben Beispielen, wie für *carpet.sav* beschrieben, enthält jedoch die tatsächlichen Einstufungen durch jeden der 10 Kunden. Die Kunden wurden gebeten, die 22 Produktprofile in der Reihenfolge ihrer Präferenzen einzustufen. Die Variablen *PREF1* bis *PREF22* enthalten die IDs der zugeordneten Profile, wie in *carpet_plan.sav* definiert.
- **catalog.sav.** Diese Datendatei enthält hypothetische monatliche Verkaufszahlen für drei Produkte, die von einem Versandhaus verkauft werden. Daten für fünf mögliche Einflussvariablen wurden ebenfalls aufgenommen.
- **catalog_seasfac.sav.** Diese Datendatei ist mit *catalog.sav* identisch, außer, dass ein Set von saisonalen Faktoren, die mithilfe der Prozedur "Saisonale Zerlegung" berechnet wurden, sowie die zugehörigen Datumsvariablen hinzugefügt wurden.
- **cellular.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, bei der es um die Bemühungen eines Mobiltelefonunternehmens geht, die Kundenabwanderung zu verringern. Scores für die Abwanderungsneigung (von 0 bis 100) werden auf die Kunden angewendet. Kunden mit einem Score von 50 oder höher streben vermutlich einen Anbieterwechsel an.

- **ceramics.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, bei der es um die Bemühungen eines Herstellers geht, der ermitteln möchte, ob ein neue, hochwertige Keramiklegierung eine größere Hitzebeständigkeit aufweist als eine Standardlegierung. Jeder Fall entspricht einem Test einer der Legierungen; die Temperatur, bei der das Keramikwälzlager versagte, wurde erfasst.
- **cereal.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, bei der es um eine Umfrage geht, bei der 880 Personen nach ihren Frühstückgewohnheiten befragt wurden. Außerdem wurden Alter, Geschlecht, Familienstand und Vorliegen bzw. Nichtvorliegen eines aktiven Lebensstils (auf der Grundlage von mindestens zwei Trainingseinheiten pro Woche) erfasst. Jeder Fall entspricht einem Teilnehmer.
- **clothing_defects.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, bei der es um die Qualitätskontrolle in einer Bekleidungsfabrik geht. Aus jeder in der Fabrik produzierten Charge entnehmen die Kontrolleure eine Stichprobe an Bekleidungsartikeln und zählen die Anzahl der Bekleidungsartikel die inakzeptabel sind.
- **coffee.sav.** Diese Datendatei enthält Daten zum wahrgenommenen Image von sechs Eiskaffeemarken ((Kennedy, Riquier, als auch Sharp, 1996)). Bei den 23 Attributen des Eiskaffee-Image sollten die Teilnehmer jeweils alle Marken auswählen, die durch dieses Attribut beschrieben werden. Die sechs Marken werden als “AA”, “BB”, “CC”, “DD”, “EE” und “FF” bezeichnet, um Vertraulichkeit zu gewährleisten.
- **contacts.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, bei der es um die Kontaktlisten einer Gruppe von Vertretern geht, die Computer an Unternehmen verkaufen. Die einzelnen Kontaktpersonen werden anhand der Abteilung, in der sie in ihrem Unternehmen arbeiten und anhand ihrer Stellung in der Unternehmenshierarchie in Kategorien eingeteilt. Außerdem werden der Betrag des letzten Verkaufs, die Zeit seit dem letzten Verkauf und die Größe des Unternehmens, in dem die Kontaktperson arbeitet, aufgezeichnet.
- **creditpromo.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, bei der es um die Bemühungen eines Kaufhauses geht, die Wirksamkeit einer kürzlich durchgeführten Kreditkarten-Werbeaktion einzuschätzen. Dazu wurden 500 Karteninhaber nach dem Zufallsprinzip ausgewählt. Die Hälfte erhielt eine Werbebeilage, die einen reduzierten Zinssatz für Einkäufe in den nächsten drei Monaten ankündigte. Die andere Hälfte erhielt eine Standard-Werbebeilage.
- **customer_dbase.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, bei der es um die Bemühungen eines Unternehmens geht, das die Informationen in seinem Data Warehouse nutzen möchte, um spezielle Angebote für Kunden zu erstellen, die mit der größten Wahrscheinlichkeit darauf ansprechen. Nach dem Zufallsprinzip wurde eine Untergruppe des Kundenstamms ausgewählt. Diese Gruppe erhielt die speziellen Angebote und die Reaktionen wurden aufgezeichnet.
- **customer_information.sav.** Eine hypothetische Datendatei mit Kundenmailingdaten wie Name und Adresse.
- **customer_subset.sav.** Eine Teilmenge von 80 Fällen aus der Datei *customer_dbase.sav*.
- **customers_model.sav.** Diese Datei enthält hypothetische Daten zu Einzelpersonen, auf die sich eine Marketingkampagne richtete. Zu diesen Daten gehören demografische Informationen, eine Übersicht über die bisherigen Einkäufe und die Angabe ob die einzelnen Personen auf die Kampagne ansprachen oder nicht. Jeder Fall entspricht einer Einzelperson.

- **customers_new.sav.** Diese Datei enthält hypothetische Daten zu Einzelpersonen, die potenzielle Kandidaten für Marketingkampagnen sind. Zu diesen Daten gehören demografische Informationen und eine Übersicht über die bisherigen Einkäufe für jede Person. Jeder Fall entspricht einer Einzelperson.
- **debate.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, die gepaarte Antworten auf eine Umfrage unter den Zuhörern einer politischen Debatte enthält (Antworten vor und nach der Debatte). Jeder Fall entspricht einem Befragten.
- **debate_aggregate.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, in der die Antworten aus *debate.sav* aggregiert wurden. Jeder Fall entspricht einer Kreuzklassifikation der bevorzugten Politiker vor und nach der Debatte.
- **demo.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, bei der es um eine Kundendatenbank geht, die zum Zwecke der Zusendung monatlicher Angebote erworben wurde. Neben verschiedenen demografischen Informationen ist erfasst, ob der Kunde auf das Angebot geantwortet hat.
- **demo_cs_1.sav.** Hierbei handelt es sich um eine hypothetische Datendatei für den ersten Schritt eines Unternehmens, das eine Datenbank mit Umfrageinformationen zusammenstellen möchte. Jeder Fall entspricht einer anderen Stadt. Außerdem sind IDs für Region, Provinz, Landkreis und Stadt erfasst.
- **demo_cs_2.sav.** Hierbei handelt es sich um eine hypothetische Datendatei für den zweiten Schritt eines Unternehmens, das eine Datenbank mit Umfrageinformationen zusammenstellen möchte. Jeder Fall entspricht einem anderen Stadtteil aus den im ersten Schritt ausgewählten Städten. Außerdem sind IDs für Region, Provinz, Landkreis, Stadt, Stadtteil und Wohneinheit erfasst. Die Informationen zur Stichprobenziehung aus den ersten beiden Stufen des Stichprobenplans sind ebenfalls enthalten.
- **demo_cs.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, die Umfrageinformationen enthält die mit einem komplexen Stichprobenplan erfasst wurden. Jeder Fall entspricht einer anderen Wohneinheit. Es sind verschiedene Informationen zum demografischen Hintergrund und zur Stichprobenziehung erfasst.
- **dmdata.sav.** Dies ist eine hypothetische Datendatei, die demografische und kaufbezogene Daten für ein Direktmarketingunternehmen enthält. *dmdata2.sav* enthält Informationen für eine Teilmenge von Kontakten, die ein Testmailing erhalten. *dmdata3.sav* enthält Informationen zu den verbleibenden Kontakten, die kein Testmailing erhalten.
- **dietstudy.sav.** Diese hypothetische Datendatei enthält die Ergebnisse einer Studie der "Stillman-Diät" (Rickman, Mitchell, Dingman, als auch Dalen, 1974). Jeder Fall entspricht einem Teilnehmer und enthält dessen Gewicht vor und nach der Diät in amerikanischen Pfund sowie mehrere Messungen des Triglyceridspiegels (in mg/100 ml).
- **dvdplayer.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, bei der es um die Entwicklung eines neuen DVD-Spielers geht. Mithilfe eines Prototyps hat das Marketing-Team Zielgruppendaten erfasst. Jeder Fall entspricht einem befragten Benutzer und enthält demografische Daten zu dem Benutzer sowie dessen Antworten auf Fragen zum Prototyp.
- **german_credit.sav.** Diese Daten sind aus dem Daten-Set "German credit" im Repository of Machine Learning Databases ((Blake als auch Merz, 1998)) an der Universität von Kalifornien in Irvine entnommen.

- **grocery_1month.sav.** Bei dieser hypothetischen Datendatei handelt es sich um die Datendatei *grocery_coupons.sav*, wobei die wöchentlichen Einkäufe zusammengefasst sind, sodass jeder Fall einem anderen Kunden entspricht. Dadurch entfallen einige der Variablen, die wöchentlichen Änderungen unterworfen waren, und der verzeichnete ausgegebene Betrag ist nun die Summe der Beträge, die in den vier Wochen der Studie ausgegeben wurden.
- **grocery_coupons.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, die Umfragedaten enthält, die von einer Lebensmittelkette erfasst wurden, die sich für die Kaufgewohnheiten ihrer Kunden interessiert. Jeder Kunde wird über vier Wochen beobachtet, und jeder Fall entspricht einer Kundenwoche und enthält Informationen zu den Geschäften, in denen der Kunde einkauft sowie zu anderen Merkmalen, beispielsweise welcher Betrag in der betreffenden Woche für Lebensmittel ausgegeben wurde.
- **guttman.sav.** Bell ((Bell, 1961)) legte eine Tabelle zur Darstellung möglicher sozialer Gruppen vor. Guttman ((Guttman, 1968)) verwendete einen Teil dieser Tabelle, bei der fünf Variablen, die Aspekte beschreiben, wie soziale Interaktion, das Gefühl der Gruppenzugehörigkeit, die physische Nähe der Mitglieder und die Formalität der Beziehung, mit sieben theoretischen sozialen Gruppen gekreuzt wurden: "crowds" (Menschenmassen, beispielsweise die Zuschauer eines Fußballspiels), "audience" (Zuhörerschaften, beispielsweise die Personen im Theater oder bei einer Vorlesung), "public" (Öffentlichkeit, beispielsweise Zeitungsleser oder Fernsehzuschauer), "mobs" (Mobs, wie Menschenmassen, jedoch mit wesentlich stärkerer Interaktion), "primary groups" (Primärgruppen, vertraulich), "secondary groups" (Sekundärgruppen, freiwillig) und "modern community" (die moderne Gesellschaft, ein lockerer Zusammenschluss, der aus einer engen physischen Nähe und dem Bedarf an spezialisierten Dienstleistungen entsteht).
- **health_funding.sav.** Hierbei handelt es sich um eine hypothetische Datei, die Daten zur Finanzierung des Gesundheitswesens (Betrag pro 100 Personen), Krankheitsraten (Rate pro 10.000 Personen der Bevölkerung) und Besuche bei medizinischen Einrichtungen/Ärzten (Rate pro 10.000 Personen der Bevölkerung) enthält. Jeder Fall entspricht einer anderen Stadt.
- **hivassay.sav.** Hierbei handelt es sich um eine hypothetische Datendatei zu den Bemühungen eines pharmazeutischen Labors, einen Schnelltest zur Erkennung von HIV-Infektionen zu entwickeln. Die Ergebnisse des Tests sind acht kräftiger werdende Rotschattierungen, wobei kräftigeren Schattierungen auf eine höhere Infektionswahrscheinlichkeit hindeuten. Bei 2.000 Blutproben, von denen die Hälfte mit HIV infiziert war, wurde ein Labortest durchgeführt.
- **hourlywagedata.sav.** Hierbei handelt es sich um eine hypothetische Datendatei zum Stundenlohn von Pflegepersonal in Praxen und Krankenhäusern mit unterschiedlich langer Berufserfahrung.
- **insurance_claims.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, bei der es um eine Versicherungsgesellschaft geht, die ein Modell zur Kennzeichnung verdächtiger, potenziell betrügerischer Ansprüche erstellen möchte. Jeder Fall entspricht einem Anspruch.
- **insure.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, bei der es um eine Versicherungsgesellschaft geht, die die Risikofaktoren untersucht, die darauf hinweisen, ob ein Kunde die Leistungen einer mit einer Laufzeit von 10 Jahren abgeschlossenen Lebensversicherung in Anspruch nehmen wird. Jeder Fall in der Datendatei entspricht einem Paar von Verträgen, je einer mit Leistungsforderung und der andere ohne, wobei die beiden Versicherungsnehmer in Alter und Geschlecht übereinstimmen.

- **judges.sav.** Hierbei handelt es sich um eine hypothetische Datendatei mit den Wertungen von ausgebildeten Kampfrichtern (sowie eines Sportliebhabers) zu 300 Kunstturnleistungen. Jede Zeile stellt eine Leistung dar; die Kampfrichter bewerteten jeweils dieselben Leistungen.
- **kinship_dat.sav.** Rosenberg und Kim ((Rosenberg als auch Kim, 1975)) haben 15 Bezeichnungen für den Verwandtschaftsgrad untersucht (Tante, Bruder, Cousin, Tochter, Vater, Enkelin, Großvater, Großmutter, Enkel, Mutter, Neffe, Nichte, Schwester, Sohn, Onkel). Die beiden Analytiker baten vier Gruppen von College-Studenten (zwei weibliche und zwei männliche Gruppen), diese Bezeichnungen auf der Grundlage der Ähnlichkeiten zu sortieren. Zwei Gruppen (eine weibliche und eine männliche Gruppe) wurden gebeten, die Bezeichnungen zweimal zu sortieren; die zweite Sortierung sollte dabei nach einem anderen Kriterium erfolgen als die erste. So wurden insgesamt sechs "Quellen" erzielt. Jede Quelle entspricht einer Ähnlichkeitsmatrix mit 15×15 Elementen. Die Anzahl der Zellen ist dabei gleich der Anzahl der Personen in einer Quelle minus der Anzahl der gemeinsamen Platzierungen der Objekte in dieser Quelle.
- **kinship_ini.sav.** Diese Datendatei enthält eine Ausgangskonfiguration für eine dreidimensionale Lösung für *kinship_dat.sav*.
- **kinship_var.sav.** Diese Datendatei enthält die unabhängigen Variablen *gender* (Geschlecht), *gener*(Generation) und *degree* (Verwandtschaftsgrad), die zur Interpretation der Dimensionen einer Lösung für *kinship_dat.sav* verwendet werden können. Insbesondere können sie verwendet werden, um den Lösungsraum auf eine lineare Kombination dieser Variablen zu beschränken.
- **marketvalues.sav.** Diese Datendatei betrifft Hausverkäufe in einem Neubaugebiet in Algonquin, Illinois, in den Jahren 1999–2000. Diese Verkäufe sind in Grundbucheinträgen dokumentiert.
- **nhis2000_subset.sav.** Die "National Health Interview Survey (NHIS)" ist eine große, bevölkerungsbezogene Umfrage in unter der US-amerikanischen Zivilbevölkerung. Es werden persönliche Interviews in einer landesweit repräsentativen Stichprobe von Haushalten durchgeführt. Für die Mitglieder jedes Haushalts werden demografische Informationen und Beobachtungen zum Gesundheitsverhalten und Gesundheitsstatus eingeholt. Diese Datendatei enthält eine Teilmenge der Informationen aus der Umfrage des Jahres 2000. National Center for Health Statistics. National Health Interview Survey, 2000. Datendatei und Dokumentation öffentlich zugänglich. *ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/Datasets/NHIS/2000/*. Zugriff erfolgte 2003.
- **ozone.sav.** Die Daten enthalten 330 Beobachtungen zu sechs meteorologischen Variablen zur Vorhersage der Ozonkonzentration aus den übrigen Variablen. Bei früheren Untersuchungen ((Breiman als auch Friedman, 1985), (Hastie als auch Tibshirani, 1990)) fanden Wissenschaftler einige Nichtlinearitäten unter diesen Variablen, die die Standardverfahren bei der Regression behindern.
- **pain_medication.sav.** Diese hypothetische Datendatei enthält die Ergebnisse eines klinischen Tests für ein entzündungshemmendes Medikament zur Schmerzbehandlung bei chronischer Arthritis. Von besonderem Interesse ist die Zeitdauer, bis die Wirkung des Medikaments einsetzt und wie es im Vergleich mit bestehenden Medikamenten abschneidet.
- **patient_los.sav.** Diese hypothetische Datendatei enthält die Behandlungsaufzeichnungen zu Patienten, die wegen des Verdachts auf Herzinfarkt in das Krankenhaus eingeliefert wurden. Jeder Fall entspricht einem Patienten und enthält diverse Variablen in Bezug auf den Krankenhausaufenthalt.

- **patlos_sample.sav.** Diese hypothetische Datendatei enthält die Behandlungsaufzeichnungen für eine Stichprobe von Patienten, denen während der Behandlung eines Herzinfarkts Thrombolytika verabreicht wurden. Jeder Fall entspricht einem Patienten und enthält diverse Variablen in Bezug auf den Krankenhausaufenthalt.
- **polishing.sav.** Hierbei handelt es sich um die Datendatei "Nambeware Polishing Times" aus der Data and Story Library. Sie bezieht sich auf die Bemühungen eines Herstellers von Metallgeschirr (Nambe Mills, Santa Fe, New Mexico) zur zeitlichen Planung seiner Produktion. Jeder Fall entspricht einem anderen Artikel in der Produktpalette. Für jeden Artikel sind Durchmesser, Polierzeit, Preis und Produkttyp erfasst.
- **poll_cs.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, bei der es um Bemühungen geht, die öffentliche Unterstützung für einen Gesetzentwurf zu ermitteln, bevor er im Parlament eingebracht wird. Die Fälle entsprechen registrierten Wählern. Für jeden Fall sind County, Gemeinde und Wohnviertel des Wählers erfasst.
- **poll_cs_sample.sav.** Diese hypothetische Datendatei enthält eine Stichprobe der in *poll_cs.sav* aufgeführten Wähler. Die Stichprobe wurde gemäß dem in der Plandatei *poll.csplan* angegebenen Stichprobenplan gezogen und in dieser Datendatei sind die Einschlusswahrscheinlichkeiten und Stichprobengewichtungen erfasst. Beachten Sie jedoch Folgendes: Da im Stichprobenplan die PPS-Methode (PPS: probability proportional to size; Wahrscheinlichkeit proportional zur Größe) verwendet wird, gibt es außerdem eine Datei mit den gemeinsamen Auswahlwahrscheinlichkeiten (*poll_jointprob.sav*). Die zusätzlichen Variablen zum demografischen Hintergrund der Wähler und ihrer Meinung zum vorgeschlagenen Gesetzentwurf wurden nach der Ziehung der Stichprobe erfasst und zur Datendatei hinzugefügt.
- **property_assess.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, in der es um die Bemühungen eines für einen Bezirk (County) zuständigen Immobilienbewerters geht, trotz eingeschränkter Ressourcen die Einschätzungen des Werts von Immobilien auf dem aktuellsten Stand zu halten. Die Fälle entsprechen den Immobilien, die im vergangenen Jahr in dem betreffenden County verkauft wurden. Jeder Fall in der Datendatei enthält die Gemeinde, in der sich die Immobilie befindet, den Bewerter, der die Immobilie besichtigt hat, die seit dieser Bewertung verstrichene Zeit, den zu diesem Zeitpunkt ermittelten Wert sowie den Verkaufswert der Immobilie.
- **property_assess_cs.sav** Hierbei handelt es sich um eine hypothetische Datendatei, in der es um die Bemühungen eines für einen US-Bundesstaat zuständigen Immobilienbewerters geht, trotz eingeschränkter Ressourcen die Einschätzungen des Werts von Immobilien auf dem aktuellsten Stand zu halten. Die Fälle entsprechen den Immobilien in dem betreffenden Bundesstaat. Jeder Fall in der Datendatei enthält das County, die Gemeinde und das Wohnviertel, in dem sich die Immobilie befindet, die seit der letzten Bewertung verstrichene Zeit sowie zu diesem Zeitpunkt ermittelten Wert.
- **property_assess_cs_sample.sav.** Diese hypothetische Datendatei enthält eine Stichprobe der in *property_assess_cs.sav* aufgeführten Immobilien. Die Stichprobe wurde gemäß dem in der Plandatei *property_assess.csplan* angegebenen Stichprobenplan gezogen und in dieser Datendatei sind die Einschlusswahrscheinlichkeiten und Stichprobengewichtungen erfasst. Die zusätzliche Variable *Current value* (Aktueller Wert) wurde nach der Ziehung der Stichprobe erfasst und zur Datendatei hinzugefügt.

- **recidivism.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, bei der es um die Bemühungen einer Strafverfolgungsbehörde geht, einen Einblick in die Rückfallraten in ihrem Zuständigkeitsbereich zu gewinnen. Jeder Fall entspricht einem frühren Straftäter und erfasst Daten zu dessen demografischen Hintergrund, einige Details zu seinem ersten Verbrechen sowie die Zeit bis zu seiner zweiten Festnahme, sofern diese innerhalb von zwei Jahren nach der ersten Festnahme erfolgte.
- **recidivism_cs_sample.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, bei der es um die Bemühungen einer Strafverfolgungsbehörde geht, einen Einblick in die Rückfallraten in ihrem Zuständigkeitsbereich zu gewinnen. Jeder Fall entspricht einem früheren Straftäter, der im Juni 2003 erstmals aus der Haft entlassen wurde, und erfasst Daten zu dessen demografischen Hintergrund, einige Details zu seinem ersten Verbrechen sowie die Daten zu seiner zweiten Festnahme, sofern diese bis Ende Juni 2006 erfolgte. Die Straftäter wurden aus per Stichprobenziehung ermittelten Polizeidirektionen ausgewählt (gemäß dem in *recidivism_cs.csplan* angegebenen Stichprobenplan). Da hierbei eine PPS-Methode (PPS: probability proportional to size; Wahrscheinlichkeit proportional zur Größe) verwendet wird, gibt es außerdem eine Datei mit den gemeinsamen Auswahlwahrscheinlichkeiten (*recidivism_cs_jointprob.sav*).
- **rfm_transactions.sav.** Eine hypothetische Datendatei mit Kauftransaktionsdaten wie Kaufdatum, gekauften Artikeln und Geldbetrag für jede Transaktion.
- **salesperformance.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, bei der es um Bewertung von zwei neuen Verkaufsschulungen geht. 60 Mitarbeiter, die in drei Gruppen unterteilt sind, erhalten jeweils eine Standardschulung. Zusätzlich erhält Gruppe 2 eine technische Schulung und Gruppe 3 eine Praxisschulung. Die einzelnen Mitarbeiter wurden am Ende der Schulung einem Test unterzogen und die erzielten Punkte wurden erfasst. Jeder Fall in der Datendatei stellt einen Lehrgangsteilnehmer dar und enthält die Gruppe, der der Lehrgangsteilnehmer zugeteilt wurde sowie die von ihm in der Prüfung erreichte Punktzahl.
- **satisf.sav.** Hierbei handelt es sich um eine hypothetische Datendatei zu einer Zufriedenheitsumfrage, die von einem Einzelhandelsunternehmen in 4 Filialen durchgeführt wurde. Insgesamt wurden 582 Kunden befragt. Jeder Fall gibt die Antworten eines einzelnen Kunden wieder.
- **screws.sav.** Diese Datendatei enthält Informationen zu den Eigenschaften von Schrauben, Bolzen, Muttern und Reißnägeln ((Hartigan, 1975)).
- **shampoo_ph.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, bei der es um die Qualitätskontrolle in einer Fabrik für Haarpflegeprodukte geht. In regelmäßigen Zeitabständen werden Messwerte von sechs separaten Ausgangschargen erhoben und ihr pH-Wert erfasst. Der Zielbereich ist 4,5–5,5.
- **ships.sav.** Ein an anderer Stelle ((McCullagh et al., 1989)) vorgestelltes und analysiertes Daten-Set bezieht sich auf die durch Wellen verursachten Schäden an Frachtschiffen. Die Vorfallshäufigkeiten können unter Angabe von Schiffstyp, Konstruktionszeitraum und Betriebszeitraum gemäß einer Poisson-Rate modelliert werden. Das Aggregat der Betriebsmonate für jede Zelle der durch die Kreuzklassifizierung der Faktoren gebildeten Tabelle gibt die Werte für die Risikoanfälligkeit an.
- **site.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, bei der es um die Bemühungen eines Unternehmens geht, neue Standorte für die betriebliche Expansion auszuwählen. Das Unternehmen beauftragte zwei Berater unabhängig voneinander mit der Bewertung der Standorte. Neben einem umfassenden Bericht gaben die Berater auch eine

zusammenfassende Wertung für jeden Standort als “good” (gut) “fair” (mittelmäßig) oder “poor” (schlecht) ab.

- **smokers.sav.** Diese Datendatei wurde aus der Umfrage “National Household Survey of Drug Abuse” aus dem Jahr 1998 abstrahiert und stellt eine Wahrscheinlichkeitsstichprobe US-amerikanischer Haushalte dar. (*http://dx.doi.org/10.3886/ICPSR02934*) Daher sollte der erste Schritt bei der Analyse dieser Datendatei darin bestehen, die Daten entsprechend den Bevölkerungstrends zu gewichten.
- **stroke_clean.sav.** Diese hypothetische Datendatei enthält den Zustand einer medizinischen Datenbank, nachdem diese mithilfe der Prozeduren in der Option “Data Preparation” bereinigt wurde.
- **stroke_invalid.sav.** Diese hypothetische Datendatei enthält den ursprünglichen Zustand einer medizinischen Datenbank, der mehrere Dateneingabefehler aufweist.
- **stroke_survival.** In dieser hypothetischen Datendatei geht es um die Überlebenszeiten von Patienten, die nach einem Rehabilitationsprogramm wegen eines ischämischen Schlaganfalls mit einer Reihe von Problemen zu kämpfen haben. Nach dem Schlaganfall werden das Auftreten von Herzinfarkt, ischämischem Schlaganfall und hämorrhagischem Schlaganfall sowie der Zeitpunkt des Ereignisses aufgezeichnet. Die Stichprobe ist auf der linken Seite abgeschnitten, da sie nur Patienten enthält, die bis zum Ende des Rehabilitationprogramms, das nach dem Schlaganfall durchgeführt wurde, überlebten.
- **stroke_valid.sav.** Diese hypothetische Datendatei enthält den Zustand einer medizinischen Datenbank, nachdem diese mithilfe der Prozedur “Daten validieren” überprüft wurde. Sie enthält immer noch potenziell anomale Fälle.
- **survey_sample.sav.** Diese Datendatei enthält Umfragedaten einschließlich demografischer Daten und verschiedener Meinungskennzahlen. Sie beruht auf einer Teilmenge der Variablen aus der NORC General Social Survey aus dem Jahr 1998. Allerdings wurden zu Demonstrationszwecken einige Daten abgeändert und weitere fiktive Variablen hinzugefügt.
- **telco.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, bei der es um die Bemühungen eines Telekommunikationsunternehmens geht, die Kundenabwanderung zu verringern. Jeder Fall entspricht einem Kunden und enthält verschiedene Informationen zum demografischen Hintergrund und zur Servicenutzung.
- **telco_extra.sav.** Diese Datendatei ähnelt der Datei *telco.sav*, allerdings wurden die Variablen “tenure” und die Log-transformierten Variablen zu den Kundenausgaben entfernt und durch standardisierte Log-transformierte Variablen ersetzt.
- **telco_missing.sav.** Diese Datendatei ist eine Untermenge der Datendatei *telco.sav*, allerdings wurde ein Teil der demografischen Datenwerte durch fehlende Werte ersetzt.
- **testmarket.sav.** Diese hypothetische Datendatei bezieht sich auf die Pläne einer Fast-Food-Kette, einen neuen Artikel in ihr Menü aufzunehmen. Es gibt drei mögliche Kampagnen zur Verkaufsförderung für das neue Produkt. Daher wird der neue Artikel in Filialen in mehreren zufällig ausgewählten Märkten eingeführt. An jedem Standort wird eine andere Form der Verkaufsförderung verwendet und die wöchentlichen Verkaufszahlen für das neue Produkt werden für die ersten vier Wochen aufgezeichnet. Jeder Fall entspricht einer Standort-Woche.
- **testmarket_1month.sav.** Bei dieser hypothetischen Datendatei handelt es sich um die Datendatei *testmarket.sav*, wobei die wöchentlichen Verkaufszahlen zusammengefasst sind, sodass jeder Fall einem Standort entspricht. Dadurch entfallen einige der Variablen, die

wöchentlichen Änderungen unterworfen waren, und die verzeichneten Verkaufszahlen sind nun die Summe der Verkaufszahlen während der vier Wochen der Studie.

- **tree_car.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, die demografische Daten sowie Daten zum Kaufpreis von Fahrzeugen enthält.
- **tree_credit.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, die demografische Daten sowie Daten zu früheren Bankkrediten enthält.
- **tree_missing_data.sav** Hierbei handelt es sich um eine hypothetische Datendatei, die demografische Daten sowie Daten zu früheren Bankkrediten enthält und eine große Anzahl fehlender Werte aufweist.
- **tree_score_car.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, die demografische Daten sowie Daten zum Kaufpreis von Fahrzeugen enthält.
- **tree_textdata.sav.** Eine einfache Datendatei mit nur zwei Variablen, die vor allem den Standardzustand von Variablen vor der Zuweisung von Messniveau und Wertelabels zeigen soll.
- **tv-survey.sav.** Hierbei handelt es sich um eine hypothetische Datendatei zu einer Studie, die von einem Fernsehstudio durchgeführt wurde, das überlegt, ob die Laufzeit eines erfolgreichen Programms verlängert werden soll. 906 Personen wurden gefragt, ob sie das Programm unter verschiedenen Bedingungen ansehen würden. Jede Zeile entspricht einem Befragten; jede Spalte entspricht einer Bedingung.
- **ulcer_recurrence.sav.** Diese Datei enthält Teilinformationen aus einer Studie zum Vergleich der Wirksamkeit zweier Therapien zur Vermeidung des Wiederauftretens von Geschwüren. Es stellt ein gutes Beispiel für intervallzensierte Daten dar und wurde an anderer Stelle ((Collett, 2003)) vorgestellt und analysiert.
- **ulcer_recurrence_recoded.sav.** In dieser Datei sind die Daten aus *ulcer_recurrence.sav* so umstrukturiert, dass das Modell der Ereigniswahrscheinlichkeit für jedes Intervall der Studie berechnet werden kann und nicht nur die Ereigniswahrscheinlichkeit am Ende der Studie. Sie wurde an anderer Stelle ((Collett et al., 2003)) vorgestellt und analysiert.
- **verd1985.sav.** Diese Datendatei enthält eine Umfrage ((Verdegaal, 1985)). Die Antworten von 15 Subjekten auf 8 Variablen wurden aufgezeichnet. Die relevanten Variablen sind in drei Sets unterteilt. Set 1 umfasst *alter* und *heirat*, Set 2 besteht aus *pet* und *news* und in Set 3 finden sich *music* und *live*. Die Variable *pet* wird mehrfach nominal skaliert und die Variable *Alter* ordinal. Alle anderen Variablen werden einzeln nominal skaliert.
- **virus.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, bei der es um die Bemühungen eines Internet-Dienstanbieters geht, der die Auswirkungen eines Virus auf seine Netzwerke ermitteln möchte. Dabei wurde vom Moment der Virusentdeckung bis zu dem Zeitpunkt, zu dem die Virusinfektion unter Kontrolle war, der (ungefähre) prozentuale Anteil infizierter E-Mail in den Netzwerken erfasst.
- **wheeze_steubenville.sav.** Hierbei handelt es sich um eine Teilmenge der Daten aus einer Langzeitstudie zu den gesundheitlichen Auswirkungen der Luftverschmutzung auf Kinder ((Ware, Dockery, Spiro III, Speizer, als auch Ferris Jr., 1984)). Die Daten enthalten wiederholte binäre Messungen des Keuchens von Kindern aus Steubenville, Ohio, im Alter

von 7, 8, 9 und 10 Jahren sowie eine unveränderlichen Angabe, ob die Mutter im ersten Jahr der Studie rauchte oder nicht.

- **workprog.sav.** Hierbei handelt es sich um eine hypothetische Datendatei zu einem Arbeitsprogramm der Regierung, das versucht, benachteiligten Personen bessere Arbeitsplätze zu verschaffen. Eine Stichprobe potenzieller Programmteilnehmer wurde beobachtet. Von diesen Personen wurden nach dem Zufallsprinzip einige für die Teilnahme an dem Programm ausgewählt. Jeder Fall entspricht einem Programmteilnehmer.

Anhang

B

Notices

Patent No. 7,023,453

The following paragraph does not apply to the United Kingdom or any other country where such provisions are inconsistent with local law: SPSS INC., AN IBM COMPANY, PROVIDES THIS PUBLICATION "AS IS" WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF NON-INFRINGEMENT, MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. Some states do not allow disclaimer of express or implied warranties in certain transactions, therefore, this statement may not apply to you.

This information could include technical inaccuracies or typographical errors. Changes are periodically made to the information herein; these changes will be incorporated in new editions of the publication. SPSS Inc. may make improvements and/or changes in the product(s) and/or the program(s) described in this publication at any time without notice.

Any references in this information to non-SPSS and non-IBM Web sites are provided for convenience only and do not in any manner serve as an endorsement of those Web sites. The materials at those Web sites are not part of the materials for this SPSS Inc. product and use of those Web sites is at your own risk.

When you send information to IBM or SPSS, you grant IBM and SPSS a nonexclusive right to use or distribute the information in any way it believes appropriate without incurring any obligation to you.

Information concerning non-SPSS products was obtained from the suppliers of those products, their published announcements or other publicly available sources. SPSS has not tested those products and cannot confirm the accuracy of performance, compatibility or any other claims related to non-SPSS products. Questions on the capabilities of non-SPSS products should be addressed to the suppliers of those products.

This information contains examples of data and reports used in daily business operations. To illustrate them as completely as possible, the examples include the names of individuals, companies, brands, and products. All of these names are fictitious and any similarity to the names and addresses used by an actual business enterprise is entirely coincidental.

COPYRIGHT LICENSE:

This information contains sample application programs in source language, which illustrate programming techniques on various operating platforms. You may copy, modify, and distribute these sample programs in any form without payment to SPSS Inc., for the purposes of developing,


Java

Bibliografie

Bell, E. H. 1961. *Social foundations of human behavior: Introduction to the study of sociology*. New York: Harper & Row.

Blake, C. L., als auch C. J. Merz. 1998. "UCI Repository of machine learning databases." Available at http://www.ics.uci.edu/~mlearn/MLRepository.html.

Breiman, L., als auch J. H. Friedman. 1985. Estimating optimal transformations for multiple regression and correlation. *Journal of the American Statistical Association*, 80, .

Collett, D. 2003. *Modelling survival data in medical research*, 2 (Hg.). Boca Raton: Chapman & Hall/CRC.

Davison, A. C., als auch D. V. Hinkley. 2006. *Bootstrap Methods and their Application*. : Cambridge University Press.

Green, P. E., als auch V. Rao. 1972. *Applied multidimensional scaling*. Hinsdale, Ill.: Dryden Press.

Green, P. E., als auch Y. Wind. 1973. *Multiattribute decisions in marketing: A measurement approach*. Hinsdale, Ill.: Dryden Press.

Guttman, L. 1968. A general nonmetric technique for finding the smallest coordinate space for configurations of points. *Psychometrika*, 33, .

Hartigan, J. A. 1975. *Clustering algorithms*. New York: John Wiley and Sons.

Hastie, T., als auch R. Tibshirani. 1990. *Generalized additive models*. London: Chapman and Hall.

Kennedy, R., C. Riquier, als auch B. Sharp. 1996. Practical applications of correspondence analysis to categorical data in market research. *Journal of Targeting, Measurement, and Analysis for Marketing*, 5, .

McCullagh, P., als auch J. A. Nelder. 1989. *Generalized Linear Models*, 2nd (Hg.). London: Chapman & Hall.

Price, R. H., als auch D. L. Bouffard. 1974. Behavioral appropriateness and situational constraints as dimensions of social behavior. *Journal of Personality and Social Psychology*, 30, .

Rickman, R., N. Mitchell, J. Dingman, als auch J. E. Dalen. 1974. Changes in serum cholesterol during the Stillman Diet. *Journal of the American Medical Association*, 228, .

Rosenberg, S., als auch M. P. Kim. 1975. The method of sorting as a data-gathering procedure in multivariate research. *Multivariate Behavioral Research*, 10, .

Shao, J., als auch D. Tu. 1995. *The Jackknife and Bootstrap*. New York: Springer.

Van der Ham, T., J. J. Meulman, D. C. Van Strien, als auch H. Van Engeland. 1997. Empirically based subgrouping of eating disorders in adolescents: A longitudinal perspective. *British Journal of Psychiatry*, 170, .

Verdegaal, R. 1985. *Meer sets analyse voor kwalitatieve gegevens (in niederländischer Sprache)*. Leiden: Department of Data Theory, Universität Leiden.

Ware, J. H., D. W. Dockery, A. Spiro III, F. E. Speizer, als auch B. G. Ferris Jr.. 1984. Passive smoking, gas cooking, and respiratory health of children living in six cities. *American Review of Respiratory Diseases*, 129, .

Index